

Santra, Tapes (2011) *Evolutionarily stable and fragile modules of yeast biochemical network*.  
PhD thesis.

<http://theses.gla.ac.uk/2644/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

# **Evolutionarily stable and fragile modules of yeast biochemical network**



University  
of Glasgow

Tapesh Santra

Department of Computing Science

University of Glasgow

Thesis submitted for the degree of Doctor of Philosophy

Supervised by: Prof. Mark A Girolami and Prof. Walter Kolch

I would like to dedicate this thesis to my loving parents ...

## **Acknowledgements**

I would like to thank both my supervisors Prof. Mark Girolami and Prof. Walter Kolch for their valuable guidance, advice and suggestions throughout the course of this study. I would also like to thank all the members of the Inference group, Ben, Dom, Gary, Lisa, Minjung, Roberta, Ronan, Simon, Tamara, Tanima, Theo and Vlad (alphabetical order) for their support, help and suggestions. I am supported by a Research Councils UK Dorothy Hodgkin Postgraduate Scholarship award.

## **Abstract**

Gene and protein interaction networks have evolved to precisely specify cell fates and functions. Here, we analyse whether the architecture of these networks affects evolvability. We find evidence to suggest that in yeast these networks are mainly acyclic, and that evolutionary changes in these parts do not affect their global dynamic properties. In contrast, feedback loops strongly influence dynamic behaviour and are often evolutionarily conserved. Feedback loops are often found to reside in a clustered manner by means of coupling and nesting with each other in the molecular interaction network of yeast. In these clusters some feedback mechanisms are biologically vital for the operation of the module and some provide auxiliary functional assistance. We find that the biologically vital feedback mechanisms are highly conserved in both transcription regulation and protein interaction network of yeast. In particular, long feedback loops and oscillating modules in protein interaction networks are found to be biologically vital and hence highly conserved. These data suggest that biochemical networks evolve differentially depending on their structure with acyclic parts being permissive to evolution while cyclic parts tend to be conserved.

# Contents

<b>Nomenclature</b>	<b>xxvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Contribution of this thesis . . . . .	6
<b>2 Global dynamic properties of biochemical networks</b>	<b>10</b>
2.1 Dynamic properties of an acyclic motif . . . . .	12
2.2 Dynamic properties of a cyclic motif . . . . .	17
2.3 The dynamic properties of a generic arbitrarily large biochemical network . . . . .	19
2.4 Arbitrarily large dynamic systems and Michaelis Menten kinetics . . . . .	21
2.5 Arbitrarily large dynamic systems and mass action kinetics	33
2.6 Interpretation of the dynamic properties of biochemical networks . . . . .	49
2.7 The implications of network dynamics on network evo- lution . . . . .	51
<b>3 Structural properties of biochemical networks</b>	<b>54</b>
3.1 Network Data Preprocessing . . . . .	60

3.2	Clustering protein protein interaction data . . . . .	62
3.3	Directionality in binary protein interaction data . . . . .	63
3.4	Relative abundance of cyclic and acyclic structures in cellular state specific biochemical networks . . . . .	65
3.5	Some critical points regarding the above analysis . . . . .	66
3.5.1	Protein interaction data . . . . .	66
3.5.2	Assigning directionality in protein interaction data	68
3.5.3	PPI clustering . . . . .	69
3.6	Structural properties and evolution of biochemical net- works . . . . .	70
<b>4</b>	<b>Differences in network topologies preferentially occur in acyclic network parts</b>	<b>76</b>
4.1	Comparison of genomes interactomes of <i>S. cerevisiae</i> and <i>S. pombe</i> . . . . .	77
4.1.1	Genome conservation in budding and fission yeast	78
4.2	Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks . . . . .	82
4.3	t <i>S. cerevisiae</i> genes which take part in biochemical feedback interactions evolve slowly . . . . .	94
4.4	Conservation of feedback mechanism in different sig- nalling and cell cycle network of yeasts . . . . .	97
<b>5</b>	<b>A comparative study of transcriptional feedback modules in different yeast species</b>	<b>107</b>
5.1	Dominant and auxiliary feedback in transcriptional feed- back modules . . . . .	110
5.1.1	Nitrogen catabolite repression (NCR) circuit . .	110

5.1.2	Dynamics of wild type NCR circuit . . . . .	113
5.1.3	DEH1 as an auxiliary gene . . . . .	120
5.1.4	Dynamics of a <i>deh1</i> $\Delta$ mutant . . . . .	122
5.1.5	Dominant and auxiliary feedback in the budding yeast NCR circuit . . . . .	123
5.1.6	Conservation of the main feedback loop of bud- ding yeast NCR module . . . . .	126
5.1.7	Carbon catabolite repression (CCR) circuit . . .	127
5.1.8	Dynamics of budding yeast CCR network . . . .	131
5.1.9	MTH1 as an auxiliary gene . . . . .	135
5.1.10	Dynamics of <i>mtl1</i> $\Delta$ mutant . . . . .	137
5.1.11	Dominant and auxiliary feedback in the CCR pathway of budding yeast . . . . .	138
5.1.12	Evolution of carbon catabolite repression circuit in different yeast species . . . . .	140
5.2	A critical assessment of the analysis presented in this chapter . . . . .	141
5.3	Concluding remarks . . . . .	143
<b>6</b>	<b>Study of feedback modules in protein interaction networks of different yeast species</b>	<b>145</b>
6.1	Pheromone response pathway . . . . .	147
6.1.1	Dynamics of yeast pheromone response pathway	150
6.1.2	Dynamics of a <i>kss1</i> $\Delta$ mutant . . . . .	153
6.1.3	Dominant and auxiliary feedback loops in pheromone response pathway . . . . .	155
6.1.4	Conservation of pheromone response pathway in other yeast species . . . . .	157



6.2	Osmoregulation pathway . . . . .	159
6.3	Dynamics of wild type budding yeast osmoregulation pathway . . . . .	160
6.3.1	The roles of ptp2,3 . . . . .	166
6.3.2	The dynamics of ptp2,3 $\Delta$ mutant . . . . .	168
6.3.3	The dominant and auxiliary feedback of yeast osmoregulation pathway . . . . .	168
6.3.4	Evolution of osmoregulation pathway in differ- ent yeast species . . . . .	170
6.4	A critical assessment and concluding remarks . . . . .	173
<b>7</b>	<b>Evolution of feedback modules which generate autonomous oscillation</b>	<b>174</b>
<b>8</b>	<b>Conclusion</b>	<b>178</b>
	<b>References</b>	<b>198</b>

# List of Figures

- 2.1 Network motifs commonly found in the transcription regulatory and protein protein interaction networks of *Saccharomyces cerevisiae*. (a) Feed forward loop, (b) Bifan motif, (c) Dense Overlapping Regulon motif, (d) Regulatory chain motif, (e) autoregulatory motif and (f) multicomponent loop motif.  $X, X_i, Y, Y_i, Z, Z_i$  represents interacting genes and proteins. The arrows represent biochemical interactions. (a), (b),(c), and (d) are acyclic motifs; (e) and (f) have feedback loops. . . . 13
- 2.2 Stability analysis of a feed forward loop. The schematic diagram of a feed forward loop is shown in figure 2.2(a) and a sample phase portrait with two trajectories are shown in figure 2.2(b). The trajectories shown in figure 2.2(b) originated from initial concentrations IC1 and IC2 and terminated at the equilibrium point EP. The tubes surrounding the trajectories can be termed as stability tubes and indicate that any state in those tubes will always remain in them. . . . . 16

2.3	Bifurcation of auto-activating feedback loop. (a) shows a schematic diagram of an auto activating feedback where $X$ represents a gene or protein which activates itself. (b) shows the surface of bifurcation in the parameter space of the equation shown in 2.3. . . . .	18
2.4	Dynamic properties of biochemical networks (BNs). An arbitrary BN is shown in (a). In (b) we show how we determine the dynamic properties of acyclic motifs. First, a mathematical model of a feed forward loop (FFL) is made. Then the Jacobian matrix of this model is calculated. The Jacobian matrix indicates monostable dynamics. In (c) we show the dynamic properties of arbitrarily large acyclic network. (d) and (e) shows dynamic properties of small and arbitrarily large feedback loops. The feedback loops have bifurcative dynamics. (f) shows the dynamic properties of an arbitrary biochemical network. Arbitrary biochemical networks are locally bifurcative in the cyclic regions. . . . .	31
2.5	A three node toy network . . . . .	32
2.6	Representation of the chemical reaction network shown in figure 2.4 in complex space. . . . .	37

- 3.1 Flowchart showing the processing and clustering of protein protein interaction data to form cellular state specific molecular interaction networks. Firstly we made a list of all the GO terms present in [Luscombe et al. \(2004b\)](#)'s dataset. We filtered in all the PPIs which are annotated with the GO terms present in the above mentioned list. Before clustering we applied [Liu et al. \(2009\)](#) 's algorithm to detect directional interactions (such as phosphorylation, dephosphorylation, acetylation etc.) and the relevant directions of signal flow, i.e. which protein phosphorylates whom etc. Once the directional interactions are identified and directions of interactions are determined we clustered the PPI data. The GO distribution of each cellular state specific subnetwork is considered as prior observations and used as initial cluster definition. The GO annotation of each PPI is then used as feature for the clustering algorithm and the filtered set of PPIs are clustered in different cellular phases. Finally the clustered PPI subnetworks are merged with their corresponding GRNs to make cellular phase specific integrated biochemical network. The fractions of proteins and DNAs of each subnetwork which take part in cyclic structures are then calculated. . . . . 57

4.1	Conservation of genomes and gene regulatory networks in budding and fission yeast. (a) shows the number of conserved open reading frames in both species and (b) shows the percentage of genetic interaction conserved in both species. . . . .	80
-----	---	----

- 4.2 Conservation of network structures in cyclic (a) and acyclic (b) parts of yeast metabolic network. Each block represents the conservation score of the cyclic (a) and acyclic structure (b) of the corresponding species. For example, the third block in the second row of figure (a) maps the conservation score  $C_{CDU-CGL}^{cyc}$  in the colour space shown on the right of each figure. Figure (a) is visibly lighter than figure (b), except from first and fifth column, suggesting, generally, higher conservation in the network structures of the cyclic parts of yeast metabolic pathways. The first and fifth column of both figure (a) and (b) suggests the opposite for *C. albicans* and *D. hanseii*. The first(fifth) column of figure (a) suggests that only around (50-60%) of the cyclic structures found in the other yeast species are also found present in *C. albicans*(*D. hanseii*). But a closer look at the first(fifth) row suggest that (80-90%) of the cyclic structures found in the *C. albicans*(*D. hanseii*) are also present in other yeast species. This can only be possible if there are fewer cyclic structures in the metabolic networks of *C. albicans* and *D. hanseii* compared to the other species. Most of the cyclic structures that are present in the metabolic networks of *C. albicans* and *D. hanseii* are also present in the other yeast species but since the other species have many more cycles in their metabolic networks the opposite is not true. The anomaly in the first and fifth column of figure (b) can be explained in a similar way. . . . . 84

- 4.3 : Conservation of enzyme sequences in the cyclic(a) and acyclic(b) parts of the metabolic networks of different yeast species. Each small square in figure (a) and (b) maps the  $Os_{species1-species2}^{cyc}$  and  $Os_{species1-species2}^{acyc}$  to the colour space shown at the right. It can be seen that each small square in figure (a) is visibly lighter than the corresponding square of figure(b) suggesting that higher percentage of enzymes which take part in feedback interactions have orthologs in other species compared to those which take part in acyclic reactions chains. . . . . 86
- 4.4 PDF of rate of evolution in the conserved enzymes(Smith-Waterman-Gotoh sequence similarity score >100) found in cyclic and acyclic parts of the metabolic networks of all 13 yeast species. The rate of evolution is represented by evolutionary distance calculated from the similarity scores and ranges from 0-1. Enzymes whose orthologs have an evolutionary distance closer to 0 evolve at a slower rate and vice versa. The above picture shows that the probability density function of the rate of evolution of the enzymes which take part in cyclic interactions has a considerably sharper peak near zero compared to that of the acyclic parts. This suggests that a larger fraction of the enzymes which take part in the cyclic reactions of the metabolic networks of different yeast species evolve in a slower rate. . . . . 89

4.5	A Bayesian distance measure between the probability distributions of $\Delta_a$ and $\Delta_c$ . The expected distances between the distributions of $\Delta_a$ and $\Delta_c$ , denoted by $\sqrt{E(\Sigma_{ac})}$ , are plotted against the bin sizes $m = 2 \dots 50$ . $\sqrt{E(\Sigma_{aa})}$ and $\sqrt{E(\Sigma_{cc})}$ are also plotted which represents the expected distance between two samples drawn from the same distribution. It can be seen from this figure that $\sqrt{E(\Sigma_{ac})} > \sqrt{E(\Sigma_{aa/acc})} \forall m = 2 \dots 50$ . . . . .	91
4.6	Quantile-quantile plot of $\Delta_a$ and $\Delta_c$ . . . . .	92
4.7	Differential evolution in the cyclic and acyclic parts of yeast metabolic network. In this figure we compare phospholipid biosynthesis pathways of <i>S. cerevisiae</i> and <i>C. albicans</i> . The feedback mechanisms responsible for the glyoxylate cycle, serine and threonine biosynthesis are conserved in both species. Main differences arise in its regulation by exogenous precursors, inositol and choline, which is governed by the acyclic parts of the pathway that is responsible for the biosynthesis of lipids. The red nodes represent the compounds which are parts of conserved cyclic modules present in the metabolic networks of both species. The acyclic parts of this pathway are represented by dark ( <i>S. cerevisiae</i> ) and light gray ( <i>C. albicans</i> ) nodes. The Kennedy pathways are not explicitly shown. . . . .	93



- 4.8 Differential evolution in the amino-acid sequences of the genes related to the feedback mechanisms of *S. cerevisiae* biochemical networks. (A) shows the Bayesian distance measures between  $\Delta^l$  and  $\Delta^t$ . (B) shows the quantile-quantile plot of  $\Delta^l$  and  $\Delta^t$  . . . . . 96
- 4.9 Conserved regulatory feedback module in the osmoregulation pathways of *S. cerevisiae* and *S. pombe*. Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey. . . . . 98
- 4.10 Conserved regulatory feedback module in the cadmium response pathways of *S. cerevisiae* and *S. pombe*. Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey. . . . . 99
- 4.11 Conserved regulatory feedbacks in the cell cycle modules of *S. cerevisiae* and *S. pombe*. Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey. CLB5 and CIG2 are shown in grey because the feedback loop mediated by them are not conserved in both the species. CLB5 mediates a positive feedback loop with SBF in *S. cerevisiae* whereas CIG2, the *S. pombe* ortholog of CLB5 mediates a negative feedback loop with MBF, the *S. pombe* ortholog of SBF. The rest of the feedback loops in this modules are conserved in exact form. . . . . 100

4.12	Conserved regulatory feedback module in the galactose metabolism pathways of <i>S. cerevisiae</i> and <i>K. lactis</i> . Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey. RAG1 is shown in grey because the feedback loop between GAL80, GAL4 and RAG1 might not be present in <i>K. lactis</i> , because, unlike GAL2, the <i>S. cerevisiae</i> homolog of RAG1, it does not act as a galactose sensor, rather it act as a glucose transporter in <i>K. lactis</i> . . . . .	101
4.13	Conserved regulatory feedback module in the drug resistance pathway of <i>S. cerevisiae</i> and <i>C. glabrata</i> . Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey. . . . .	102
5.1	Cyclic modules reside in the hierarchical middle layer of yeast GRNs. The module shown in panel (a) is part of the yeast sporulation GRN. The modules shown in panel (b) and (c) are part of stress response GRNs. The cyclic modules are drawn in red nodes and blue edges. The subnetworks are generated from regulatory interaction data <a href="#">Lee et al. (2002)</a> ; <a href="#">Luscombe et al. (2004b)</a> ; <a href="#">Yu and Gerstein (2006)</a> . The regulatory hierarchy is made by imposing a partial order on the set of genes which form these subnetworks. We used only TF-DNA interaction data to build these subnetworks in order to simplify the visualization procedure. . . . .	108

5.2	Wild type nitrogen catabolite repression circuit of budding yeast. Arrows indicate activation and blunt arrows indicate repression. The feedback module in this network is formed due to transcription regulations of GAT1, DAL80 and DEH1 genes by each other. . . . .	112
5.3	Simulation of the model shown in equation 5.1 (a) and the observed expression levels (b) of components of the NCR circuit during nitrogen starvation. . . . .	116
5.4	Bifurcation diagrams for the wild type NCR model shown in equation 5.1. The control mechanism of the NCR circuit relies on nitrogen state dependent control of its degradation rates (Boczko et al. (2005)).The observed saddle node bifurcations(SN) are mainly due to fluctuations in degradation rates ( $\beta, \phi, \eta, \rho, \mu, \chi$ , see equation 5.1). Interestingly, the concentration changes of DAL80 caused by bifurcation are negligible, whereas DEH1 goes through a stiff hysteresis curve which switches its state between 'on' ( $> 0.5$ ) and 'off' ( $< 0.5$ ). Hence, in the wild type NCR circuit DEH1 acts as a nitrogen state dependent toggle switch. . . . .	118

5.5	Responses of the wild type NCR circuit due to changes in initial concentrations of DAL80 and DEH1 mRNAs. We have simulated the responses of the wild type NCR circuit for four different combinations of initial concentrations, i.e. ([DAL80] =0, [DEH1]=0), ([DAL80] =0, [DEH1]=1) , ([DAL80] =1, [DEH1]=0) and ([DAL80] =1, [DEH1]=1). It can be seen from this figure that in steady state DAL80 always comes down to 'off' state. However, the steady state response of DEH1 gene depends on the initial state of DAL80 and follows the Boolean logic $[DEH1]^{n+1} = \overline{[DAL80]^n}$ which realizes a D type flip flop (truth table shown in table 5.2). The importance of such a backup system lies in the fact that the functioning of DAL80 is mainly transient and always comes down to its 'off' state needing a memory to remember its initial condition for future reference. . . .	119
5.6	Schematic diagram of a NCR circuit in a <i>deh1</i> Δ mutant.	121
5.7	Bifurcation diagrams of the <i>deh1</i> Δ mutant NCR model. No hysteresis occurs due to perturbations in the reaction rates of the <i>deh1</i> Δ mutant NCR model. Hence, it does not produce bistability. For the convenience of visualisation we have shown the homeostasis levels GAT1 and DAL80 only. . . . .	124

- 5.8 The 'dominant' and 'auxiliary' feedback loops of budding yeast NCR circuit. The thick red arrows indicate the interactions that forms the main feedback mechanism of budding yeast NCR circuit. The auxiliary feedback loops are represented by black dotted lines. . . . . 125
- 5.9 Evolution of GATA factor genes in fungi. Same colours indicate homologous genes. For example Gat1 in *S. cerevisiae*, *C. albicans*, and *P. stipitis*, and KLLA0F25300g in *K. lactis* are homologous and so on. The domain organisation is shown in a box right to the gene names and colour coded. For example, GAT1 and GZF are written in same colour. The species of Saccharomycetaceae family have at least one homolog of GZF(GATA binding Zn finger) and one of GZF-LZ(GATA binding Zn finger with an additional Leucine Zipper). However, species of other families contain double GZF-GZF encoding genes or many single GZF genes without any LZ. Hence, it is predicted that the ancestral yeast had only one GZF encoding genes and the other combinations appeared later in evolution. The data shown in this figure is collected from **Rossa and Peter (2005)** and **Low and Atchly (2000)**. 128

- 5.10 The glucose repression circuit of budding yeast ([Kaniak et al. \(2004\)](#)). The transcription regulatory feedback mechanisms of fig 5.10(a) are shown in figure 5.10(b). In figure 5.10(b), the arrowheads represent activation and the blunt arrows represent repression. The black solid lines represent direct transcription regulation and the blue dashed lines represent indirect transcription regulation via chains of protein interactions. In this case, both indirect transcription regulations are glucose dependent. . . . . 130
- 5.11 Simulation of the mathematical model of carbon catabolite repression circuit as shown in equation 5.4. 5.11(a) shows the mRNA levels normalised to unit average as found by Tu et. al. [Tu et al. \(2005a\)](#). 5.11(a) is generated by SCEPTRANS [Kudlicki et al. \(2007\)](#). 5.11(b) shows the result of simulation of our model. . . . . 133

5.12	Bifurcation analysis of the dynamic behavior of the <i>S. cerevisiae</i> CCR circuit (only MTH1 is shown) due to perturbations in its parameters (see equation 5.4, table 5.3). Both saddle node(SN) and Hopf bifurcation(HB) are observed indicating that it can show bistable behaviour and oscillations. The genes in CCR network can exhibit sustained oscillations between 'on' and 'off' states due to changes in its parameters, some of which are glucose dependent (e.g., $k_{21}$ , $k_{22}$ ). The maximum and minimum concentrations during oscillation are shown by dashed lines. The oscillatory regions in the bifurcation diagrams are highlighted in yellow . . . . .	134
5.13	The variations in period of oscillation (T) due to changes in parameter values. The period (T) can vary from 60 mins to 350 mins depending on the parameter values. . .	136
5.14	The transcriptional feedback mechanism of a <i>mtl1Δ</i> mutant. The <i>mtl1Δ</i> mutant has only one feedback loop made of SNF3 and MIG1/2 . . . . .	137
5.15	Bifurcation analysis of <i>mtl1Δ</i> mutant CCR network of budding yeast. No bifurcation occurs for a wide range of parameter values in the neighbourhood of the operating point. Hence the <i>mtl1Δ</i> mutant is unable to produce sustained oscillation. . . . .	139

- 5.16 Glucose repression pathway of *Candida albicans* (Jeffrey and Brown (2009)). In this figure we have shown only part of the pathway which forms the transcriptional feedback mechanism. Some of the other interactions shown by Jeffrey and Brown (2009) are not shown in this figure for the convenience of visualisation. . . . . 141
- 5.17 Biologically vital or functionally 'dominant' feedback loops in a transcriptional feedback module are most likely to be conserved. (a) A conceptual diagram of evolutionary conservation in a four gene feedback module. The functionally dominant loop is drawn as solid line and remains conserved during evolution. Auxiliary loops are draw as thin broken lines and are less conserved. (b) shows the preferential conservation of the possibly 'dominant' loops in a cell cycle related transcriptional feedback modules of *S. cerevisiae* and *S. pombe*. (c) shows another example of evolutionary conservation of the GAL80-GAL4-GAL1/3 feedback module in *S. cerevisiae* and *K. lactis*. The GAL2 mediated feedback in the same module of *S. cerevisiae* is not conserved in *K. lactis*. (d) shows the conservation of functionally 'dominant' feedback loops of the carbon catabolite repression module in *S. cerevisiae* and *C. albicans*. The MTH1-MIG1/2 mediated auxiliary feedback loop is not conserved in *C. albicans*. . . . . 144



6.1	Schematic diagram of pheromone response pathway of budding yeast. This pathway has a long MAPK cascade which activates Bar1 and degrades pheromone resulting in a feedback regulation in pheromone sensing. The long feedback encloses many small transcriptional and proteomic feedback loops. Here we have shown only the feedback loops caused by protein interaction. Such as the positive feedback loop which arises due to the activation of Fus3 by Ste12 and vice versa. The self regulatory transcriptional feedback of Ste12 is not shown in this figure. . . . .	149
6.2	Bifurcation analysis of the mathematical model (equation 6.1, table 6.1) of the budding yeast pheromone response pathway. The model exhibits hysteresis and bistability (saddle node bifurcations, denoted by SN in the figure) due to fluctuations in many of its parameters including the basal rate of pheromone production $B_\alpha$ . The results are similar to Paliwal et al. (2007). . . . .	154
6.3	Bifurcation analysis of a $kss1\Delta$ mutant. In a $kss1\Delta$ mutant the hysteresis is either absent or too narrow to make a difference. . . . .	156

6.4	Pheromone response pathways in budding (a) and fission yeast (b). Boxes of same colours represent homologous genes. For example, mam2/map3, ste2/ste3; $G_\alpha$ (budding yeast), $G_\alpha$ (fission yeast); byr2, ste11; byr1, ste7; spk1, fus3/kss1 are homologous <a href="#">Herskowitz (1995)</a> . Similarly, ste12(budding yeast), ste11(fission yeast); rgs1, sst2 <a href="#">Hughes et al. (1994)</a> ; <a href="#">Vohra et al. (2003)</a> ; and sxa2, bar1; <a href="#">Imai and Yamamoto (1992)</a> shk1, ste20 <a href="#">Marcus et al. (1995)</a> are homologous. . . . .	158
6.5	The osmoregulation pathway of budding yeast. . . . .	161
6.6	The response of the mathematical model of the osmoregulation pathway. The concentration of the end product glycerol of the osmoregulation pathway as observed by <a href="#">Krantz et al. (2004b)</a> is plotted against the glycerol concentration simulated by our model. . . . .	164
6.7	The simulation results of osmoregulation pathway of yeast. (a) shows the concentration levels observed by <a href="#">Krantz et al. (2004b)</a> and (b) shown the simulation results of our model. . . . .	165
6.8	Bifurcation analysis of the wild type budding yeast osmoregulation pathway. No bifurcation occurs due to changes in the parameters of the osmoregulation pathway of budding yeast. . . . .	167

- 6.9 Bifurcation diagrams of the *ptp2,3Δ* mutant osmoregulation model. The model shows both saddle node(SN) which promotes bistability and Hopf bifurcation (HB) which promotes oscillation. The oscillatory regions are highlighted in yellow. . . . . 169
- 6.10 Conservation of the osmoregulation pathway. Schematic diagrams of the osmoregulation pathways of budding(a) and fission yeast (b) and *Candida lusitanae* (c). Boxes of same color represent homologous genes, i.e., *mcs4*, *ssk1*; *wak1*, *ssk2,22*; *wis1*,*pbs2*; *sty1*,*hog1* and *pyp2*,*ptp2* are homologous Cottarel (1997). Similarly, *mak1,2,3*, *sln1*; *mpr1*,*ypd1* are homologous Santos and Shiozaki (2001). Some of interactions of the osmoregulation pathways are taken from Aiba et al. (1998); Ikner and Shiozaki (2005). (c)The osmoregulation pathway of *Candida lusitanae* as described by Boissard et al. (2008). . 171

6.11 In protein interaction networks long range feedback modules are more likely to be conserved compared to the small nested ones. (A) shows a conceptual diagram of evolutionary conservation of long feedback loops. The small nested feedback loops are drawn in gray and are less likely to be conserved in different yeast species. (B) and (C) shows the evolutionary conservation of pheromone response pathways in <i>S. cerevisiae</i> and <i>S. pombe</i> . All the long range feedback loops of <i>S. cerevisiae</i> pheromone response pathway are conserved in <i>S. pombe</i> . Only the small Kss1 mediated small feedback loop is not present in <i>S. pombe</i> . (D), (E) and (F) shows the evolutionary conservation of the osmoregulation pathway in <i>S. cerevisiae</i> , <i>S. pombe</i> and <i>C. lusitanae</i> . The Sln1-Pbs2-Hog1 mediated long feedback loop is found conserved in all three species, however the small Ptp2/3 mediated negative feedback loops are conserved in <i>S. cerevisiae</i> and <i>S. pombe</i> but not in <i>C. lusitanae</i> (Aiba et al. (1998); Boissard et al. (2008); Cottarel (1997); Ikner and Shiozaki (2005); Santos and Shiozaki (2001)). . . . .	172
---	-----

- 7.1 Conservation of a prototypical cell cycle oscillator in budding (a) and fission (b) yeast. The interactions of the budding yeast network are from [Chen et al. \(2000\)](#), and of the fission yeast network from [Sveiczer et al. \(2004\)](#). Boxes of the same colour represent homologous genes. The prototypical yeast cell cycle module is conserved in both yeasts. A rather simplistic description of the prototypical cell cycle oscillator is as follows. Cyclins, Clb2, Clb5(Cdc13,Cig2) are inhibited by Sic1(Rum1) at the beginning(G1 phase) of *S. cerevisiae* (*S. pombe*) cell cycle. As the cell grows Cln2/3(not shown in the figure, similar mechanism in *S.pombe* is not yet clear) degrades Sic1(Rum1) and release Clb2(Cdc13) which represses Sic1(Rum1) thus letting free Clb2, Clb5(Cig2,Cdc13) accumulate in the cell which triggers the G1-S transition. In middle and late phases of the cell cycle, Clb2(Cdc13) activates Cdc20(Slp1), which activates Hct1(Ste9). Activated Hct1(Ste9) degrades Clb2(Cdc13) releasing Sic1 from its inhibitory effect and the cell returns to G1 resembling an oscillator. . . . . 177

# Chapter 1

## Introduction

### 1.1 Background

How evolution works at the molecular level is a fascinating field of investigation. The study of evolution at the level of proteins and DNA was commenced in the 1960s to understand the evolution of enzyme functions, as was the use of nucleic acid divergence as molecular clock to study species divergence, and to study the origin of nonfunctional and junk DNA (**Graur and Li (2000)**). Recent advances in genomics such as gene sequencing, high throughput protein characterisation, bioinformatics and systems biology have prompted a dramatic increase in interest in this field. Recent efforts to unravel the mechanisms of molecular evolution involve comparison of sequenced genomes of different species. Comparing gene and protein sequences of different organisms reveals

important information about the origin, conservation and divergence of different genes and proteins found in them. Examples of such studies are [Butler et al. \(2009\)](#); [Kellis et al. \(2003\)](#); [Peng et al. \(2005\)](#). These studies reveal that some genes are better conserved among certain species compared to others. The reasons behind such disparities in conservation among different genes are still unclear despite several efforts. Some of the preliminary efforts to answer the above question focussed on the structural and functional properties of individual genes and proteins. The basic assumption of these studies was that the structure and function of individual genes and proteins are somehow related to their evolutionary potential.

For example, [Lipman et al. \(2002\)](#) suggested that the evolutionary conservation of a protein depends on its sequence length, proteins with longer amino acid sequences are better conserved than the ones with relatively short sequences. [Fraser et al. \(2003\)](#) suggested that the proteins which take part in larger number of biochemical interactions are better conserved, which agrees with the findings of [Lipman et al. \(2002\)](#) since proteins with long amino acid sequences are more likely to be involved in large number of protein interactions due to presence of abundant potential interaction domains. However, [Agrafioti et al. \(2005\)](#) suggested that the proteins which are highly expressed are better conserved than the others which are relatively less abundant. On the contrary, [Warringer and Blomberg \(2006\)](#) found that differentially expressed proteins with very different sequence lengths may evolve at the same rate. [Hahn et al. \(2004\)](#); [Jordan et al. \(2003\)](#) opposed the findings of [Fraser et al. \(2003\)](#) by claiming that there is no dependence between protein evolution rate and

the number of protein interactions. [Bloom and Adami \(2003\)](#) claimed that the apparent interdependence between protein interaction numbers and the evolution rate found by [Fraser et al. \(2003\)](#) is a result of biases in the protein interaction data used in the study.

The conclusions of the above research works are contradictory. The confusion grew further as high throughput data started to become available which enabled scientists to compare the functions of conserved genes and proteins among different species. For example, [Rustici et al. \(2007\)](#) compared the set of genes which responds to changing iron and copper levels in *S. cerevisiae* and *S. pombe*. They found that only a small set of genes with conserved sequences perform the same function in both species, whereas a rather large set of conserved genes have different function in these species. This finding is in sharp contradiction to the very assumption that the conservation of genes and proteins are somehow related to their individual structure and function. In another study [Peng et al. \(2005\)](#) compared the cell cycle regulated genes in both *S. cerevisiae* and *S. pombe* and found that a significant number of homologous genes are regulated differently in these species. This finding is in agreement with the conclusion of [Rustici et al. \(2007\)](#) that conserved genes may have different functions in different species. Further experiments by [Dixon et al. \(2008\)](#) suggested that only 23-30% of genetic interactions are conserved between *S. cerevisiae* and *S. pombe* whereas they share almost 60% of the conserved genome([Peng et al. \(2005\)](#)). Since the interaction patterns are conserved to a much lesser extent than the gene sequences themselves it can be inferred that a significant number of conserved genes operate differently in these species.



The above findings suggest the possibility of an alternative explanation of molecular evolution. Such an explanation may be possible as the focus of functional genomics shifts from the functional analysis of individual genes and proteins towards the functional analysis of gene and protein interaction networks. But the enormous complexity of these network structures has so far prevented scientists from determining the functional properties of the genetic and proteomic interaction networks of even the most well studied organisms such as *S. cerevisiae*. Since the architecture of these networks are complex it seemed necessary to understand the architectural properties of these networks in order to understand their dynamics. Some of the first efforts of structural characterisation of large scale biochemical networks <sup>1</sup> revealed some interesting characteristics of these networks. For example, [Yook et al. \(2004\)](#) suggested that biochemical networks have scale free architectures, [Pržulj et al. \(2004\)](#) suggested that these networks have the same properties as geometric random graphs, [Milo et al. \(2002\)](#); [Yeger-Lotem et al. \(2004\)](#) found that some specific small network modules, 'network motifs' as termed by them, are enriched in biochemical networks compared to random networks. [Pržulj et al. \(2006\)](#) further integrated the previous methods to characterise biochemical networks by the degree distribution of 'graphlets'. However, the correlation between molecular evolution and the architectural properties of biochemical networks remained unclear.

On the other hand, in a parallel effort, biologists and anthropologists such as [Davidson and Erwin \(2006\)](#); [Kwon and Cho \(2007\)](#) found that

---

<sup>1</sup>from here on we shall refer to both gene regulatory and protein interaction networks as biochemical networks unless otherwise stated

certain small subnetworks of interacting genes are conserved for millions of years. It has been argued that these gene regulatory network (GRN) components are preserved for millions of years because of their pivotal role in designing animal body plan (Davidson and Erwin (2006)). These components are termed as "kernel" components. Interestingly, despite the rigid behaviour of the kernel components against evolutionary changes the network linkages surrounding them have altered almost beyond recognition since pre-Cambrian time. The mechanisms which account for such disparity in the rate of evolution in different parts of GRN is still unclear.

After the discovery of network motifs, the statistically over-represented small sub-networks found in GRNs, by Milo et al. (2002), Kashtan and Alon (2005) argued that the reason behind their over abundance is the preserving of their particular interaction patterns during evolution. However, this theory has some drawbacks. First of all, the argument of Kashtan and Alon (2005) does not explain the presence of "kernel" components as discovered by Davidson and Erwin (2006), secondly, Mazurie et al. (2005) presented strong and appealing evidence against the theory of Kashtan and Alon (2005). Other noticeable efforts to elucidate this matter made by the mathematical and statistical community, such as those of Ciliberti et al. (2007) and Krishnan et al. (2008), are mainly focussed on very specific structural and functional properties of GRNs which either helps new innovations to take place during evolution, or emerges from beneficial evolution. To the best of our knowledge, very few efforts have been made so far to answer the question as to how exactly evolutionary mechanisms alter the architecture of an organism's

biochemical network. To be more precise, which parts of a GRN are most likely to be altered and which parts are most likely to be conserved during the course of evolution? This calls for an integrative analysis of both structural and functional genomics to understand the role of how dynamic properties of biochemical networks relate to their evolution.

## 1.2 Contribution of this thesis

We have analysed molecular interaction data along with publicly available microarray and protein activation data obtained from a model eukaryot, *Saccharomyces cerevisiae*, commonly known as baker's yeast, in order to answer the questions mentioned in the previous section. In particular, we seek to identify the structural and functional properties of different parts of yeast GRN (i) which are flexible towards changes in their architecture by means of evolution, and (ii) which are rather rigid against any structural change.

The salient contributions of this thesis to the conceptual advancement of the field are twofold. Firstly, we developed a mathematical analysis to predict the dynamical behaviour of arbitrarily large biochemical networks. Usually, biochemical networks are modelled using a set of differential equations and their functionalities are predicted by simulating these equations. However, modelling the dynamic behaviour of a biochemical network at a genomic scale is not feasible due to lack of data

on interaction parameters. In this study we developed a mathematical analysis to determine the global dynamic behaviour of an arbitrarily large biochemical network from its structural properties rather than having to model each of its interactions individually. Our theoretical analysis can be used in a wide range of scientific and technological studies where complex and large nonlinear systems are common. In this thesis we show an example of applying this theory to investigate whether there is a connection between network topologies and evolution. While such a connection is plausible, neither its role and nor its nature has been systematically analysed until now.

Secondly, our results reveal an unexpected link between network design features and evolution, which in some cases contradicts many early studies on molecular evolution but explains many recent findings. We provide a concrete mathematical explanation to suggest that there is a balance between robustness of design and potential for evolution and provide evidence from available biological data to support our hypothesis. Parts of this thesis have been presented at the following conferences and workshops:

- Gene expression profiling using Dirichlet process based non parametric trans-dimensional clustering ,April 2007, Mathematical and Statistical Aspects of Molecular Biology, Manchester, U.K.
- Relationship between structure and dynamics of gene regulatory networks, March 2008, Learning in Computational System's Biology, University of Glasgow, Glasgow, U.K.

- An analysis of the relation between structure and dynamics of biochemical networks, March 2008, Mathematical and Statistical Aspects of Molecular Biology, University of Glasgow, Glasgow, U.K.
- Dynamic properties of large graphs, September 2008, Complex Networks across the Natural and Technological Sciences, University of Strathclyde, Glasgow, U.K.
- The architecture of a gene Regulatory Network is related to evolutionary potential, April 2009, Mathematical and Statistical Aspects of Molecular Biology, London, U.K.

The remainder of the thesis is organised as follows. In chapter 2 I discuss the dynamic properties of biochemical networks and the relation of these properties with molecular evolution. I generalise the theory of nonlinear dynamic systems for arbitrarily large biochemical networks and find that their acyclic and cyclic parts contribute differently to their dynamics. Further analysis led me to the question how much of a typical yeast biochemical network is made of acyclic structure and how much is made of cyclic loops. In chapter 3 I look into the structural properties of biochemical networks to answer the above question. In chapter 4 I look for evidences to support our conclusions of chapter 2 and 3 by comparing gene sequence and network structures of multiple yeast species. Chapter 5, 6 and 7 are dedicated to case studies of different network modules of yeast biochemical networks. In these case studies I analysed the dynamic properties of biochemical modules related to signalling, mating, metabolism and cell cycle and looked for their evolution in different yeast species. In chapter 8 I concluded by discussing about the work and future

directions.

## Chapter 2

# Global dynamic properties of biochemical networks

The description of “kernel” components in the gene regulatory networks of mammals by [Davidson and Erwin \(2006\)](#) raised some interesting possibilities. Since “kernels” are small subnetworks of interacting genes and proteins it is possible that the function of a “kernel” as a module is more important than the function of its individual components. Small GRN modules are also found conserved in arthropods ([Damen \(2007\)](#)), nematodes ([Ge et al. \(2006\)](#)) and echinoderms ([Hinman et al. \(2009\)](#)) due to their important roles in developmental biology. It is important to understand the contribution of different modules of a biochemical network to its global dynamic behaviour in order to determine which modules are functionally important and which are not. Hence, in this chapter we provide a mathematical analysis to determine the global

---

dynamic behaviour of biochemical networks and the contributions of different types of network modules to its operation.

Determining the global dynamic behaviour of an arbitrarily large biochemical network by conventional means is hardly possible in the current context. Firstly, these networks are extremely large in size making the realisation of an ordinary differential equation model to simulate their dynamic behaviour impossible. Secondly, the exact architecture of these networks are unknown for even the most well studied model organisms like *S. cerevisiae*. Finally, the reaction parameters for most of the biochemical interactions which take place in these networks are also unknown. Our analysis aims to determine the dynamic properties of biochemical networks from their structural properties rather than having to model all of their interactions individually.

One of the most important structural properties of biochemical networks is their modular architecture. The concept of modularity in biochemical networks was introduced by Milo et al. (2002). According to Milo et al. (2002), the modular architecture of biochemical networks is caused by overabundance of certain small network modules, called 'network motifs'. Considerable efforts have been made to explain the overabundance of network motifs in biochemical networks. Milo et al. (2002) reasoned that each network motif is an important functional module and according to Mangan et al. (2006) is functionally autonomous.

Since network motifs are functionally important small subnetworks



found in biochemical networks we decided to start with a network motif analysis of biochemical networks of *S. cerevisiae*. We chose *S. cerevisiae* because of the wide availability of interaction data. We used the same method and data used by Yeger-Lotem et al. (2004) in our analysis. We enlisted all the network motifs found in that analysis (see fig. 2 for some examples) and developed mathematical models for each of them to describe their dynamic behaviour. We followed the same method as Mangan and Alon (2003) to develop these mathematical models. The network motifs of a biochemical network can be divided into two categories cyclic and acyclic motifs. In the next two sections we shall discuss the dynamic properties of one acyclic and one cyclic motif in detail.

### 2.1 Dynamic properties of an acyclic motif

Acyclic motifs are small subnetworks of biochemical networks which do not have feedback loops. In this section we determine the dynamic properties of an acyclic motif called feed forward loop (figure 2.1(a)). Following Mangan and Alon (2003) we developed a mathematical model for a feed forward loop motif using Michaelis Menten kinetics. Michaelis Menten kinetics is a derivative of mass action law and is usually used to model regulatory interactions in GRNs. Protein interactions however are usually modelled using mass action law as they are not enzymatically catalysed. For the time being we shall stick to Michaelis Menten kinetics to derive the basic concepts of our analysis. A more generalised version

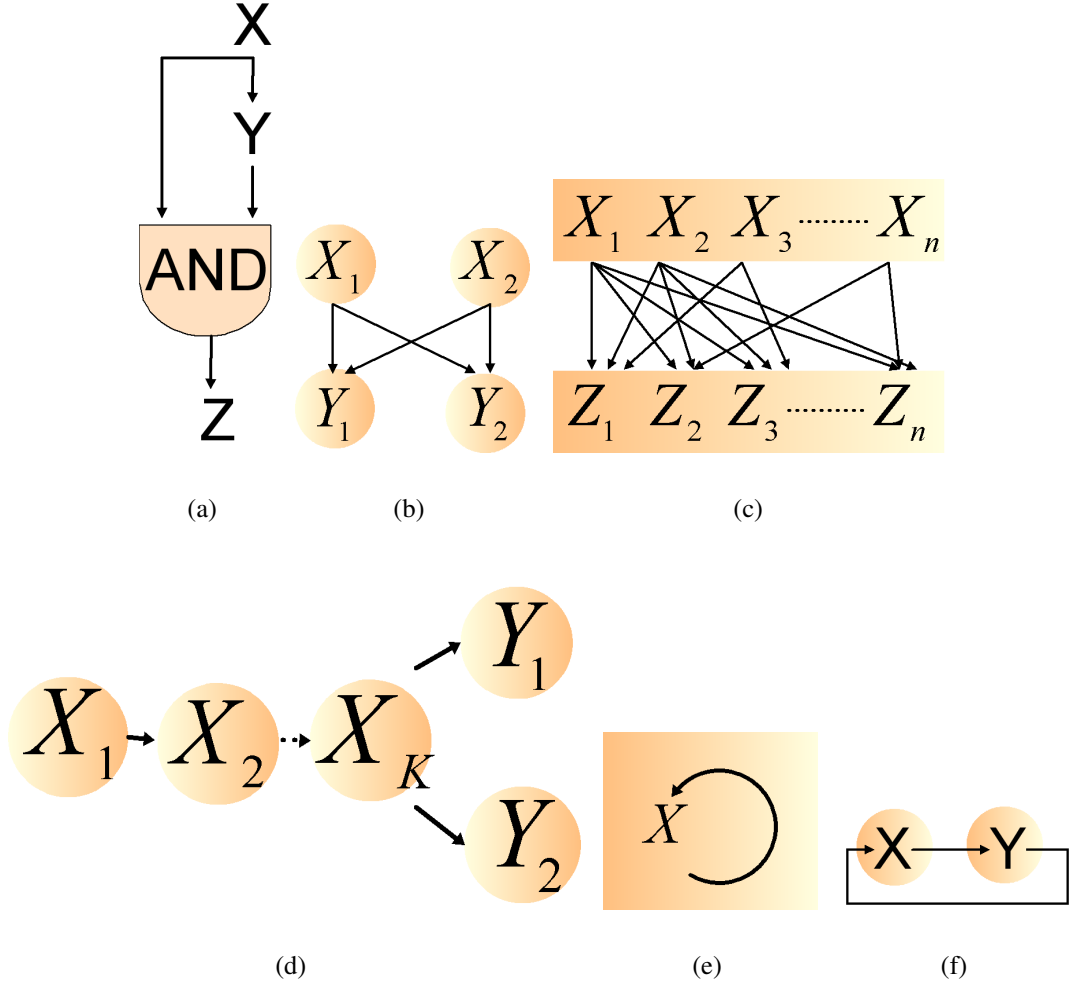


Figure 2.1: Network motifs commonly found in the transcription regulatory and protein protein interaction networks of *Saccharomyces cerevisiae*. (a) Feed forward loop, (b) Bifan motif, (c) Dense Overlapping Regulon motif, (d) Regulatory chain motif, (e) autoregulatory motif and (f) multicomponent loop motif.  $X, X_i, Y, Y_i, Z, Z_i$  represents interacting genes and proteins. The arrows represent biochemical interactions. (a), (b), (c), and (d) are acyclic motifs; (e) and (f) have feedback loops.

of our theory shall be discussed later in this chapter, Our mathematical model is shown in equation 2.1.

$$\frac{dx}{dt} = B_x - \alpha_x x$$

$$\begin{aligned}\frac{dy}{dt} &= B_y + \beta_y \left( \frac{x^2}{K_{xy}^2 + x^2} \right) - \alpha_y y \\ \frac{dz}{dt} &= B_z + \beta_z \left( \frac{x^2}{K_{xz}^2 + x^2} \right) \left( \frac{1}{1 + y^2/K_{yz}^2} \right) - \alpha_z z\end{aligned}\quad (2.1)$$

In equation 2.1,  $x, y, z$  represent the concentrations of the mRNA produced by hypothetical genes X, Y and Z.  $B_x$ ,  $B_y$  and  $B_z$  are the basal levels and  $\alpha_x$ ,  $\alpha_y$  and  $\alpha_z$  are the decay rates of  $x$ ,  $y$  and  $z$  respectively.  $\beta_y$  is the rate of activation of the gene Y by gene X and  $\beta_z$  is the rate of activation of the gene Z by both X and Y.  $K_{xy}$ ,  $K_{xz}$  and  $K_{yz}$  are Michaelis constants. The Hill coefficient is fixed at  $n = 2$  following Mangan and Alon (2003). However, the results of our analysis are independent of the value of  $n$  as will be shown later. Instead of analysing the transient and steady state response of the above module due to different input signals as done in most of the previous studies (Mangan and Alon (2003)) we analysed the stability of its steady states. Unlike the analysis of Mangan and Alon (2003) our analysis reveals the nature of the trajectories of the states of the relevant biochemical entities in response to all possible values of the kinetic parameters and initial conditions. To understand the nature of its steady states, i.e. whether they are stable, unstable or oscillatory, we computed the Jacobian matrix (shown in equation 2.2) for the set of differential equations shown in equation 2.1.

$$\mathcal{J} = \begin{pmatrix} -\alpha_x & 0 & 0 \\ \frac{2\beta_y x_{ep} K_{xy}^2}{(K_{xy}^2 + x_{ep}^2)^2} & -\alpha_y & 0 \\ \frac{2\beta_z K_{xz}^2 K_{yz}^2 x_{ep}}{M} & -\frac{2\beta_z K_{yz}^2 x_{ep}^2 y_{ep}}{M} & -\alpha_z \end{pmatrix}$$

$$M = (x_{ep}^2 + K_{xz}^2) (y_{ep}^2 + K_{yz}^2)^2 \quad (2.2)$$

The Jacobian matrix  $\mathcal{J}$  shown in equation 2.2 is lower triangular with negative real numbers in its diagonal elements. Hence the eigenvalues  $\lambda_i : i = 1, 2, 3$  of  $\mathcal{J}$  are  $\lambda_i = -\alpha_i$ . As all the eigenvalues of the Jacobian matrix  $\mathcal{J}$  are negative real numbers, i.e.,  $\lambda_i < 0$ , the system is stable near its steady state  $x_{ep}, y_{ep}, z_{ep}$  and the stability of the system does not depend on any of its parameters, i.e., the module will always remain stable for all biologically plausible values of its parameters ( $>0$ ). As the equation 2.1 has only one steady state, we conclude that a feed forward loop is a monostable system whose stability is independent of its parameter values, i.e. the reaction rate constants. A phase space representation of the equation 2.1 is shown in figure 2.2.

Interestingly, similar analysis for all possible acyclic network motifs results in the same conclusion. In other words, the steady states of all acyclic network motifs are always stable disregarding their parameter values. Further investigation reveals that the reason behind such phenomena is the acyclic nature of these network structures. The acyclicity of these motifs gives rise to triangular Jacobian matrices for the system of equations describing their dynamics. Since the dynamics of these motifs are modelled using the Michaelis Menten kinetics which essentially follow the rules of chemical kinetics, the diagonal elements of the Jacobian matrices are always negative real numbers which indicate absolute stability of their steady states. The biological implications of the above results are discussed in detail in section 2.5. However, these results are

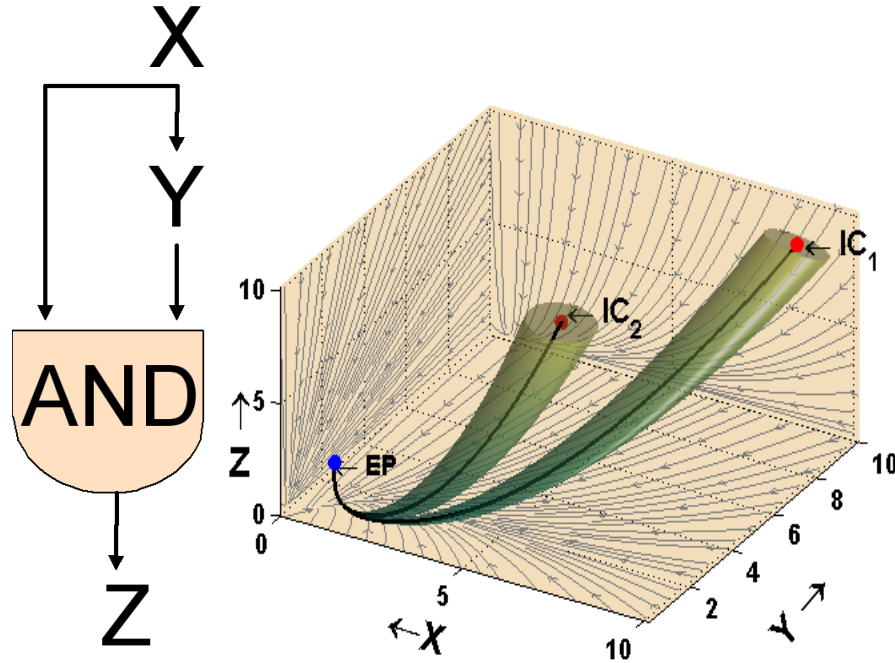


Figure 2.2: Stability analysis of a feed forward loop. The schematic diagram of a feed forward loop is shown in figure 2.2(a) and a sample phase portrait with two trajectories are shown in figure 2.2(b). The trajectories shown in figure 2.2(b) originated from initial concentrations IC1 and IC2 and terminated at the equilibrium point EP. The tubes surrounding the trajectories can be termed as stability tubes and indicate that any state in those tubes will always remain in them.

remarkable in the sense that it tells us that the stability of network motifs depend on their structural property, i.e. acyclicity, rather than interaction parameters and initial concentrations. The question that naturally occurs at this point is what kind of dynamic behaviour does a cyclic motif have. Hence in the next section we shall analyse the dynamics of a motif with a feedback loop.

## 2.2 Dynamic properties of a cyclic motif

The simplest possible feedback loop is the autoregulatory loop where a biochemical entity such as a gene or protein regulates its own expression. Hence we developed a mathematical model of a positive autoregulatory motif and analysed its dynamics. The mathematical model of a positive autoregulatory circuit is shown in equation 2.3.

$$\frac{dx}{dt} = \beta \frac{(x/K_x)^2}{1 + (x/K_x)^2} - \alpha x \quad (2.3)$$

Here,  $\beta$  is the activation rate,  $\alpha$  is the rate of decay and  $K_x$  is Michaelis constant. The auto-activating entity is monostable for  $\beta/\alpha \leq 2K_x$  and bistable for  $\beta/\alpha > 2K_x$ . The bifurcation surface in the parameter space is shown in figure 2.3(b)

The dynamic behaviour of multi-component feedback loop motifs are

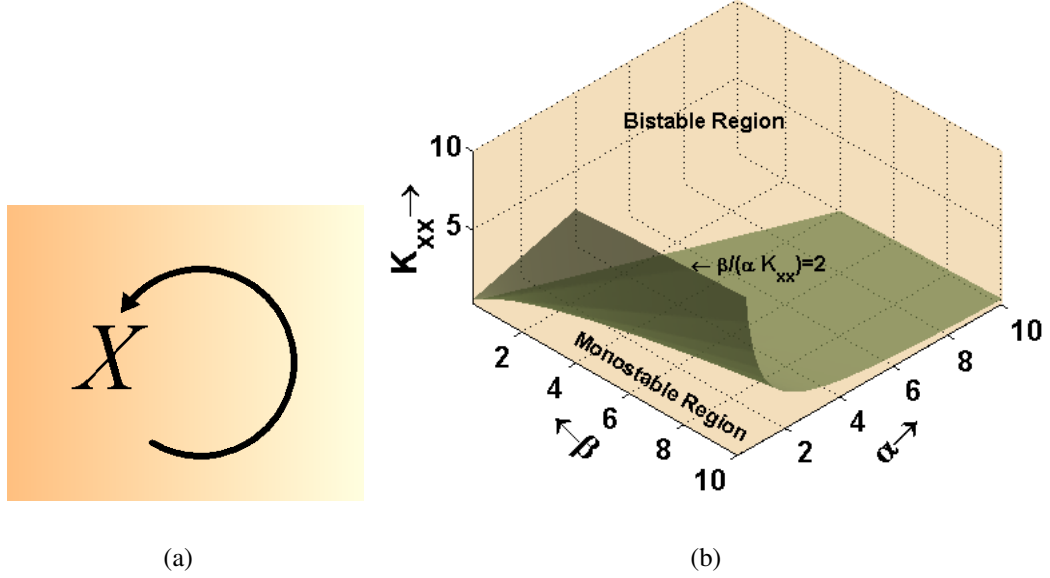


Figure 2.3: Bifurcation of auto-activating feedback loop. (a) shows a schematic diagram of an auto activating feedback where  $X$  represents a gene or protein which activates itself. (b) shows the surface of bifurcation in the parameter space of the equation shown in 2.3.

discussed in detail by [Kim et al. \(2007\)](#). The dynamics of a biochemical feedback loop can bifurcate between monostable, bistable, multi-stable, graded response and oscillatory modes depending on the initial mRNA or protein concentrations and corresponding reaction rate constants. Bistability results in toggle switch like behaviour and flip flop like memory system, multistable modules can act as genetic multiplexers, oscillators act as biological clocks and some negative feedback mechanism act as noise smoothers ([Kim et al. \(2007\)](#)). Hence, we conclude that the feedback loop motifs have rich dynamic behaviour with important potential functional roles.

*In summary, we divide the network motifs in two subcategories, cyclic*

### **2.3 The dynamic properties of a generic arbitrarily large biochemical network**

*and acyclic motifs. All acyclic motifs are stable systems and exhibit rather simple dynamic behaviour whose stability is independent of their corresponding parameters. On the other hand, cyclic motifs exhibit rich and bifurcative dynamic behaviour and may have the potential of acting as important functional modules.*

Though, at this point we have a basic idea of the autonomous functioning of some of the network motifs our goal is to find out how they shape the dynamic behaviour of a large biochemical network. In the next section we shall develop a mathematical theory to determine the dynamic properties of a generic biochemical network from the ideas gathered from the above analysis. Because of our observation that the acyclic motifs are always stable, we first generalised the theory of nonlinear dynamical systems to analyse the dynamics of an arbitrarily large acyclic biochemical network. Then we extended our theory to determine the dynamic properties of generic biochemical networks. Our mathematical analysis of the dynamic properties of biochemical networks is as follows.

### **2.3 The dynamic properties of a generic arbitrarily large biochemical network**

The dynamics of biochemical networks are represented in state space, where the change in the current state of a node (i.e. the concentrations of mRNAs, proteins etc.) depends on the states of adjacent nodes and is



### 2.3 The dynamic properties of a generic arbitrarily large biochemical network

modelled using differential equations. Appropriate functions describing biochemical interactions are rather controversial topic. Though mass action law and Michaelis Menten kinetic models, a derivative of mass action law, are among the most commonly used approaches to model biochemical interactions, there are significant controversies about their applicability in an intracellular environment. Mass action law was developed to model interactions between reactants in a relatively dilute pH buffered aqueous solution. In intracellular environments bound particles may be prevented from dissociation by their surroundings or diffusion is slow and anomalous and the model of mass action does not always describe the reaction kinetics accurately. A consensus is not yet reached to modify mass action law in order to make it applicable in intercellular environment. An alternative approach is to replace the rate constants with functions of time (commonly known as fractal kinetics, [Kopelman \(1988\)](#); [Savageau \(1995\)](#) ) and/or concentrations. In another theory, [Grima and Schnell \(2006\)](#) suggested that mass action law is applicable in an intracellular environment under certain conditions but in different rates than would be found in dilutes. However, the most widely accepted way to model the kinetics of biochemical networks is to assume that different types of chemical reaction follow different rate laws. For example, homo and hetero dimer and polymer formations and compound degradations are best described by mass action kinetics, but most of the other interactions such as phosphorylation, enzymatic reactions and gene regulations are best described by Michaelis Menten kinetics. In a biochemical network all these different types of interactions can take place at the same time. However, determining the dynamic properties of a biochemical network which contains reactions of heterogeneous kinetic

laws are not easy. Hence, we pursue this problem in the classical way. In this chapter, we shall first develop a theory to analyse large dynamic systems which follow Michaelis Menten kinetics and determine the relation between their structural and dynamical properties. Then we discuss and extend some of the existing theories which deals with the same problem for arbitrarily large dynamic systems made of mass action type reactions only. Based on these two analysis we make some conclusions about the dynamic properties of biochemical networks. We further discuss the relation between different dynamic properties of biochemical networks and their evolution.

## 2.4 Arbitrarily large dynamic systems and Michaelis Menten kinetics

Michaelis Menten kinetics is generally used to model enzymatic reactions, transcription regulatory interactions and protein phosphorylation reactions. Some other biochemical interactions such as protein complex formations are usually modelled using mass action law. However in the first part of our theory we shall assume that all the biochemical interaction that take place are of Michaelis Menten type. Such an assumption reduces the mathematical complexity of the problem to a great extent and make the proofs easily comprehensible. The first part of our analysis is divided into two theorems. The first theorem(Theorem 1) tells us that the steady states of an arbitrarily large acyclic dynamic system which

follows Michaelis Menten kinetics are always stable and their stability is independent of the reaction parameters, initial concentrations, even structural changes until and unless such changes alter their acyclic nature. The second theorem discusses the effect of feedback loops in the dynamics of such system. It tells us that an arbitrarily large generic dynamic system can have many different modes of dynamic behaviour such as monostability (one single stable steady state), multistability (multiple stable and unstable steady states), graded response, stable, unstable, periodic and aperiodic oscillation etc. But how these networks will behave is principally determined by their feedback loops with the acyclic parts having minimal role. The proofs of the above theorems are as follows.

The signal flow in a biochemical network can be represented as a directed graph. The interacting entities of the graph are represented as vertices and the directions of signal flow are represented as directional edges. We represent a biochemical network  $S$  as a triplet  $S = \langle V, E, F \rangle$ ; where  $V$  is set of vertices  $V = v_i : i = 1 \dots N$  or 'species' as commonly referred in systems biology literature. Each 'species' or vertex represents a biochemical compound such as a protein, mRNA etc.  $E$  is a set of edges  $E = \{e_{ij} = (v_i, v_j) : v_i, v_j \in V\}$  whose elements are ordered sets such that if  $(v_i, v_j) \in E$  then  $v_i \succ v_j$  where  $\succ$  denotes upstream. Each edge  $e_{ij} \in E$  represents a biochemical reaction in which  $v_i$  regulates  $v_j$ . Hence,  $v_i$  is the regulator and  $v_j$  is the regulated species.  $\succ$  defines a relation on the vertices of the graph. For linguistic convenience we call  $v_i$  the substrate and  $v_j$  the product and we assume that biochemical signals flow from substrates to products.  $F$  is a vector valued function  $F : R^N \rightarrow R^N$  such that  $\dot{\vec{x}} = F(\vec{x})$ , where  $\vec{x} = \{x_i : i = 1 \dots N\}^T$

represents the states of the vertex set  $V$ , i.e.  $F$  determines the rate of changes in the concentrations of the entities involved. In a biochemical network,  $x_i$  represents the concentration of  $v_i$ ,  $x_i = [v_i]$ . The  $i^{th}$  projection of  $F$ ,  $proj^i : F = f_i$  defines the state trajectory of the  $i^{th}$  species  $v_i$ , i.e.,  $\dot{x}_i = f_i(\vec{x})$ . In other words,  $f_i$  determines the rate of change in the concentration  $x_i$  of the  $i^{th}$  species  $v_i$ .

Furthermore, we introduce a second relation  $\Rightarrow$  on the set of the vertices. We say  $v_i \Rightarrow v_j$  if there exist a set of vertices  $V_s \subseteq V$  such that  $v_i \succ v_s^1 \succ v_s^2 \succ \dots v_s^k \succ v_j | V_s = \{v_s^l : l = 1 \dots k\}$ , i.e.  $v_j$  is reachable from  $v_i$ .  $\Rightarrow$  is 'transitive' since if  $v_i \Rightarrow v_j$  and  $v_j \Rightarrow v_k$  then  $v_i \Rightarrow v_k$ . The relation  $\Rightarrow$  has specific properties in specific type of networks. For example, in an acyclic network  $\Rightarrow$  is also irreflexive and antisymmetric. An acyclic network  $S_a$  can be defined as  $S_a = \langle V_a, E_a, F_a \rangle$  such that  $\nexists V_{sc} \subseteq V_a | v_i \succ v_{sc}^1 \succ v_{sc}^2 \dots v_{sc}^k \succ v_i$  where  $V_{sc} = \{v_{sc}^l : l = 1 \dots k\}$ , i.e. no path of finite length exists from a vertex to itself. Furthermore,  $v_i \not\succ v_i$ , i.e. no vertex regulates itself. Therefore, for acyclic networks  $v_i \not\Rightarrow v_i \forall v_i \in V_a$  which gives rise to the irreflexive property of  $\Rightarrow$ . Additionally, if  $\exists v_i, v_j \in V_a, i \neq j$  such that  $v_i \Rightarrow v_j$  and  $v_j \Rightarrow v_i$  then due to the transitive property of  $\Rightarrow$ ,  $v_i \Rightarrow v_i$ . This is a contradiction to the definition of the acyclic networks, which justifies the antisymmetric property of  $\Rightarrow$  in acyclic networks. As mentioned before, we shall discuss the dynamic properties of an arbitrarily large acyclic biochemical network first. But before going into the dynamics of acyclic biochemical networks we shall discuss some useful structural properties of directed acyclic networks in terms of definitions, propositions and lemmas. Propositions 1 and 2 and definition

1 are taken from [Thulasiraman and Swamy \(1992\)](#) and the rest of the propositions, lemmas, definitions and theorems presented in this section are contributions of this thesis.

**Proposition 1.** *A directed acyclic network with finite number of vertices(nodes) has at least one vertex(node) with indegree zero and one vertex with outdegree zero [Thulasiraman and Swamy \(1992\)](#).*

**Proof:** For a given acyclic network  $S_a = \langle V_a, E_a \rangle$ , choose any vertex  $v \in V$  and follow any path from  $v$  along the direction of the edges. We cannot return to any vertex since the network is acyclic and we must eventually terminate since  $|V|$  is finite. The only way we can get stuck is if we hit a vertex with outdegree zero. If we do the same as above, but this time, travelling in the opposite direction of the edges we shall eventually get stuck at the nodes with indegree zero.

Proposition 1 is important since it tells us that there exists at least one start vertex and one end vertex in a directed acyclic graph. We shall use this property to impose a partial order on the vertices of such graphs. To prove that such partial order exists we shall use common graph theoretic concepts such as transitive reduction and node depth, and introduce some new concepts such as depth class and some lemmas and propositions regarding depth classes

**Definition 1.** *The transitive reduction of a directed acyclic graph  $S_a = \langle V_a, E_a \rangle$  is the minimal representation of the graph  $S_a$  such that the reachability relation  $\Rightarrow$  is preserved in the reduced graph  $S_{am} = \langle V_a, E_{am} \rangle$ , i.e. if in  $S_a$ ,  $\exists v_i, v_j, v_k$  such that  $v_i \Rightarrow v_j$  and  $v_j \Rightarrow v_k$  means*

$v_i \Rightarrow v_k$  then the  $v_i, v_j, v_k$  in  $S_{am}$  has the same relation *Thulasiraman and Swamy (1992)*.

**Proposition 2.** *If  $V_r \subset V_a$  is the set of vertices with indegree zero and outdegree  $> 0$  in a directed acyclic network  $S_a$  then  $\forall v_{ai} \in V_a \setminus V_r, \exists v_r \in V_r$  such that  $v_r \Rightarrow v_a$ . In other words, every node of a directed acyclic graph can be reached from at least one of its root nodes *Thulasiraman and Swamy (1992)*.*

**Proof:** The proof is similar to the proof of proposition 1.

**Definition 2.** *We redefine the depth of a node in this thesis for our convenience. The depth of a node  $v_a$  in an acyclic graph  $S_a = \langle V_a, E_a \rangle$  is defined by the maximum number of edges one needs to travel to reach  $v_a$  from a root node  $v_r \in V_r \subset V_a$  when travelling along the edges of its transitive reduction. If  $|V_r| > 1$  and  $v_a$  is reachable from more than one root nodes, i.e.  $\exists v_{r1}, v_{r2} \dots v_{rm} \in V_r | v_{r1} \Rightarrow v_a, v_{r2} \Rightarrow v_a \dots v_{rm} \Rightarrow v_a$  and the maximum numbers of edges one needs to travel to reach  $v_a$  from  $v_{r1}, v_{r2} \dots v_{rm}$  are  $d_{r1}, d_{r2} \dots d_{rm}$  then the depth of  $v_a$  is defined as  $d_a = \max(d_{r1}, d_{r2}, \dots d_{rm})$ .*

Under this definition every node has a unique depth given a transitive reduction.

**Definition 3.** *In this definition we introduce the concept of depth classes. The set of depth classes of a directed acyclic network  $S_a$  is a partition  $\Pi_{V_a} = \{\Pi_i : i = 1, 2 \dots p\}$  on its vertex set  $V_a$  such that, if  $v_{a1}, v_{a2} \in \Pi_i$  then  $d_{a1} = d_{a2}$  where  $d_{a1}$  and  $d_{a2}$  are the depths of the vertices  $v_{a1}$  and*

$v_{a2}$ . In other words, each element of  $\Pi_{V_a}$  is a set of vertices with same depth.

For example, the set of root nodes  $V_r$  is an element of  $\Pi_{V_a}$  since all the vertices (nodes) which belong to  $V_r$  has a depth zero. Each element of the partition  $\Pi_{V_a}$  has a depth associated with it. If we denote the depth associated with  $\Pi_i \in \Pi_{V_a}$  as  $d_{\pi_i}$  then we say  $\Pi_i \triangleright \Pi_j$  if  $d_{\pi_i} > d_{\pi_j}$ .  $\triangleright$  is a relation on the elements of the depth class of a directed acyclic network. It is transitive, irreflexive and antisymmetric. Using  $\triangleright$  we can impose a full order on the elements of  $\Pi_{V_a}$ . Let us rearrange the elements of  $\Pi_{V_a}$  in ascending order such that  $\Pi_{V_a} = \{\Pi_1, \Pi_2 \dots \Pi_k\}$  and  $\Pi_{i+1} \triangleright \Pi_i \forall i \in \{1 \dots k-1\}$ .

**Proposition 3.** *If  $\Pi_i, \Pi_j \in \Pi_{V_a} | \Pi_i \triangleright \Pi_j$  then  $\nexists v_{\Pi_i} \in \Pi_i | v_{\Pi_i} \Rightarrow v_{\Pi_j} \in \Pi_j$ . In other words, a node which belongs to a depth class of depth  $d_{\pi_j}$  cannot be reached from a node which belongs to a depth class of depth  $d_{\pi_i}$  if  $d_{\pi_i} > d_{\pi_j}$ .*

**Proof:** Let us assume that there exists vertices of acyclic graph  $S_a$ ,  $v_{\Pi_i} \in \Pi_i$  and  $v_{\Pi_j} \in \Pi_j$  such that  $v_{\Pi_i} \Rightarrow v_{\Pi_j}$  where  $d_{\pi_i} > d_{\pi_j}$ . Hence, by definition of depth,  $d_{\pi_j} > d_{\pi_i}$  since  $d_{\pi_j}$  is the maximum possible depth of  $v_{\Pi_j}$ . This is a contradiction.

**Proposition 4.**  *$\nexists v_{\pi_{i1}}, v_{\pi_{i2}} \in \Pi_i \in \Pi_{V_a} | v_{\pi_{i1}} \Rightarrow v_{\pi_{i2}}$ . In other words, no two vertices of the same depth class can be reached from one another.*

**Proof:** The proof of this proposition is similar to the proof of

proposition 4. Assume the contrary, i.e.  $\exists v_{\pi_{i1}}, v_{\pi_{i2}} \in \Pi_i \in \Pi_{V_a} | v_{\pi_{i1}} \Rightarrow v_{\pi_{i2}}$ . In that case from the definition of depth of a node(vertex),  $d_{\pi_{i1}}$  must be less than  $d_{\pi_{i2}}$ . But, since,  $v_{\pi_{i1}}$  and  $v_{\pi_{i2}}$  belong to the same depth class they must have the same depth. This is a contradiction.

From propositions 3 and 4 it is obvious that if  $v_i, v_j \in V_a | v_i \Rightarrow v_j$  then  $\Pi_i \triangleright \Pi_j$  where  $v_i \in \Pi_i, v_j \in \Pi_j$ . The above observation allows us to impose a partial order on  $V_a$ .

**Proposition 5.** *It is possible to impose a partial order on the vertices of a directed acyclic graph.*

**Proof :** A partial order on the vertices of a directed acyclic graph can be imposed in many different ways. Here we propose one such method. First, we determine the transitive reduction of a given directed acyclic network  $S_a = \langle V_a, E_a \rangle$ . The definition of transitive reduction is given below. Let us consider the ordered set of depth classes  $\Pi_{V_a} = \{\Pi_i, i = 1 \dots p\} | d_{\pi_{i+1}} > d_{\pi_i} \forall i \in \{1 \dots p-1\}$ . Furthermore the sizes(cardinalities) of each depth class  $\Pi_i$  is given by  $\sigma_i$ . Relabel the elements  $\Pi_1$  as  $n_i : i = 1 \dots \sigma_i$  at random. Then relabel the elements of  $\Pi_2$  as  $n_i : i = \sigma_1 + 1 \dots \sigma_1 + \sigma_2$  at random and so on. Under this scheme of relabelling a vertex with label  $n_i$  can never be reached from a vertex with label  $n_j$  if  $j > i$ . Hence the new labelling scheme imposes a partial order on  $V_a$ . For notational convenience, from now on we shall denote each node by their new label for the rest of the first part of the proof.

**Lemma 1.** *The Jacobian matrix of an acyclic dynamic system is lower triangular with negative real diagonal elements.*



## 2.4 Arbitrarily large dynamic systems and Michaelis Menten kinetics

**Proof:** Let us consider an acyclic dynamic system  $S_a = \langle N_a, E_a, F_a \rangle$  where  $N_a$  is the set of nodes labelled under the partial sorting scheme mentioned above,  $E_a$  are the set of edges and  $F_a$  is a vector valued function such that  $F : R^{N_a} \rightarrow R^{N_a}$ .  $F_a$  represents the rates of changes of the states of the nodes of  $S_a$ . In a biochemical network,  $F_a$  represents the rates of changes of the concentrations of biochemical compounds over time. Hence,  $\frac{\partial \vec{x}}{\partial t} = F$ , where  $\vec{x} = \{x_i : i = 1 \dots N\}$  represents the states of the nodes  $n_i : i = 1 \dots N$  where  $N = |N_a|$ . The projection of  $F$  on the  $i^{th}$  dimension of the space spanned by the states of nodes is denoted by  $f_i$  and can be given by  $\frac{\partial x_i}{\partial t} = f_i$ . If the dynamics of the above system is modelled using Michaelis Menten kinetics then  $f_i$  has the form  $f_i = g(X_{ai}) - \alpha_i x_i |X_{ai} \subset X_a$ , where  $X_a$  represents the states of  $N_a$  and  $\forall n_j \in N_{ai} n_j \Rightarrow n_i$  and  $\alpha_i$  is the degradation constant for  $x_i$ . According to proposition 3 and 4, since  $n_j \Rightarrow n_i, \nexists \Pi_l \in \Pi_{N_a} | n_j, n_i \in \Pi_l$  and  $\nexists \Pi_i, \Pi_j \in \Pi_{N_a}$  such that  $n_i \in \Pi_i, n_j \in \Pi_j$  and  $\Pi_j \triangleright \Pi_i$ . If  $n_i \in \Pi_i, n_j \in \Pi_j$  and  $n_j \Rightarrow n_i$  then  $\Pi_i \triangleright \Pi_j$ , which indicates that  $i > j$ . Hence,  $\frac{\partial f_i}{\partial x_j} = 0 \forall j > i$ . On the other hand, since  $S_a$  is acyclic,  $n_i \Rightarrow n_i$  is always false. Hence,  $\frac{\partial f_i}{\partial x_i} = -\alpha_i$ . Hence the Jacobian matrix of an acyclic network  $S_a$  has the following form.

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} & \cdots & \frac{\partial f_3}{\partial x_N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_N}{\partial x_1} & \frac{\partial f_N}{\partial x_2} & \frac{\partial f_N}{\partial x_3} & \cdots & \frac{\partial f_N}{\partial x_N} \end{pmatrix} = \begin{pmatrix} -\alpha_1 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & 0 & \cdots & 0 \\ \frac{\delta f_i}{\delta x_1} & \cdots & -\alpha_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \frac{\delta f_N}{\delta x_i} & \cdots & -\alpha_N \end{pmatrix}$$

Since  $\alpha_i > 0 : i = 1 \dots N$ ,  $J$  is a lower triangular matrix with negative real diagonal elements.

**Theorem 1.** *All possible equilibria (steady states) of an acyclic biochemical network are always stable.*

**Proof** From Lemma 1,  $J$  is lower triangular with negative real diagonal elements. Its eigenvalues are the solution of the characteristics equation  $|J - I\lambda| = 0$ , where  $I$  is an identity matrix. In this case, since  $J$  is lower triangular its eigenvalues are the same as its diagonal elements  $\lambda_i = -\alpha_i : i = 1 \dots N$ . Since,  $\alpha_i$  is always positive, the eigenvalues of  $J$  are always negative real numbers which indicates absolute stability.

**Theorem 2:** Feedback mechanisms are solely responsible for the nature of stability of the equilibria (steady states) of a biochemical networks.

**Proof of Theorem 2:** We introduce a feedback loop from the  $k + 1^{th}$  vertex  $n_{k+1} \in N_a$  of  $S_a$  to the  $k^{th}$  one,  $n_k \in N_a$ , where  $n_k \Rightarrow n_{k+1}$ . The Jacobian matrix  $J$  of the new system becomes block triangular because  $\frac{\partial f_k}{\partial x_{k+1}} \neq 0$  where the rest of the matrix remains unchanged. Hence, the eigenvalues of  $J$  becomes  $\lambda_i = \frac{\partial f_i}{\partial x_i} \leq 0; i = 1..N; i \neq k, k + 1$ , whereas  $\lambda_k, \lambda_{k+1}$  are the eigenvalues of the matrix shown in equation 2.4, which essentially is the diagonal block appearing due to the feedback loop

introduced in the graph.

$$\begin{pmatrix} \frac{\partial f_k}{\partial x_k} & \frac{\partial f_k}{\partial x_{k+1}} \\ \frac{\partial f_{k+1}}{\partial x_k} & -\frac{\partial f_{k+1}}{\partial x_{k+1}} \end{pmatrix} \quad (2.4)$$

The introduction of the feedback loop does not alter the eigenvalues related to the acyclic parts of the network and they remain negative real numbers. The only possible eigenvalues which can have nonnegative real parts are those related to the feedback mechanism. Hence, the nature of the equilibrium, i.e. whether it is stable, unstable or encircled by a limit cycle is determined solely by the feedback loops. This result can be generalised for an arbitrary number of feedback loops in a biochemical network.

In the above analysis the law of biochemical interaction is assumed to follow the Michaelis Menten law. The entire procedure of determining the dynamic behaviour of a biochemical networks is shown in figure 2.4. The main convenience of assuming Michaelis Menten kinetics is that the dynamics of the nodes of the same depth are independent of each other which is not the case when using a mass action law. This phenomena is discussed in detail in the rather simple example shown below.

Let us assume a three node acyclic network,  $S_3 = \langle V_3, E_3, F_3 \rangle$ , where  $V_3 = \{v_1, v_2, v_3\}$ ,  $E_3 = \{(v_1, v_3), (v_2, v_3)\}$ , the projections of  $F_3$  on  $v_1, v_2, v_3$  are  $f_1, f_2, f_3$ . Let us assume  $v_1$  and  $v_2$  repress  $v_3$ . A graphical representation of  $S_3$  is shown in figure 2.4. Hence,  $f_1 = B_1 - \alpha_1 x_1$ ,  $f_2 = B_2 - \alpha_2 x_2$ ,  $f_3 = B_3 + \frac{k_1}{1+(x_1/\kappa_1)^n+(x_2/\kappa_2)^n} - \alpha_1 x_1$ , where  $x_i : i = 1 \dots 3$  represent the states of  $v_i : i = 1 \dots 3$ . Here,  $B_i, \alpha_i : i = 1, 2, 3$ ,

## 2.4 Arbitrarily large dynamic systems and Michaelis Menten kinetics

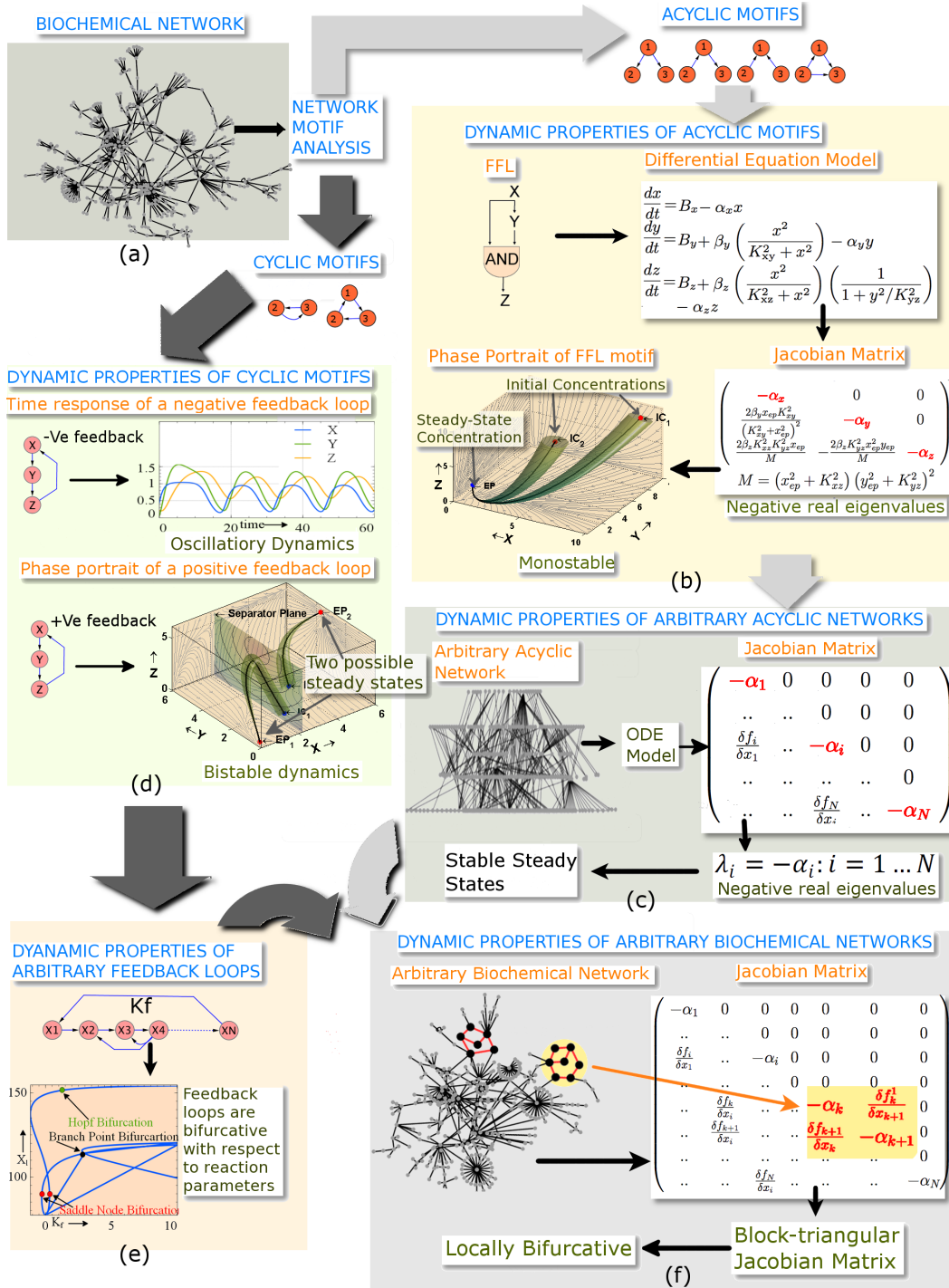


Figure 2.4: Dynamic properties of biochemical networks (BNs). An arbitrary BN is shown in (a). In (b) we show how we determine the dynamic properties of acyclic motifs. First, a mathematical model of a feed forward loop (FFL) is made. Then the Jacobian matrix of this model is calculated. The Jacobian matrix indicates monostable dynamics. In (c) we show the dynamic properties of arbitrarily large acyclic network. (d) and (e) show dynamic properties of small and arbitrarily large feedback loops. The feedback loops have bifurcative dynamics. (f) shows the dynamic properties of an arbitrary biochemical network. Arbitrary biochemical networks are locally bifurcative in the cyclic regions.

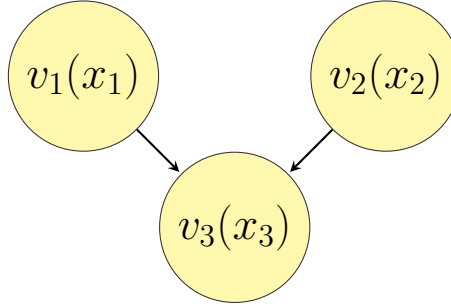


Figure 2.5: A three node toy network

$k_1, \kappa_i : i = 1, 2$  and  $n$  are all constants. In this case, both  $v_1$  and  $v_2$  have depth zero and  $\frac{\partial f_1}{\partial x_2} = 0, \frac{\partial f_2}{\partial x_1} = 0$ . Hence the Jacobian matrix for the system of differential equations  $\dot{\vec{x}} = F_3$  is lower triangular with negative real diagonal elements and the above theory holds.

Now, consider the same network shown in figure 2.4. But instead of assuming that  $v_1$  and  $v_2$  regulates  $v_3$ , let us assume that  $v_1, v_2$  interact with each other and produces  $v_3$ . The chemical reaction for such a system is given as following.



Such reaction is usually modelled using a mass action law. Under the mass action law the definition of  $F_3$  is different from the one shown above. In that case, the new definitions are  $f_1 = f_2 = -f_3 = r = -k_1 x_1 x_2$ . Here,  $k_1$  is the rate constant for reaction  $r$ . However, if we also model the degradation of biochemical compounds as suggested by [Cooper \(2006\)](#) [Steiger and Parker \(2002\)](#) and [Kornitzer \(2002\)](#) then under the new definition of  $F_3$ ,  $f_1 = r - \alpha_1 x_1, f_2 = r - \alpha_2 x_2, f_3 = -r - \alpha_3 x_3$ . Both  $v_1$  and  $v_2$  have depth zero, and  $\frac{\partial f_1}{\partial x_2} = -k_1 x_1$  is nonzero if  $x_1 > 0$ .

Since,  $\frac{\partial f_1}{\partial x_2}$  is an off-diagonal term in the Jacobian matrix of the system equation  $\dot{\vec{x}} = F_3$ , the Jacobian matrix is no longer triangular as can be seen from equation 2.7. Hence the eigenvalues of such matrices may not be guaranteed to be negative. However, in this particular case the eigenvalues of the Jacobian matrix  $J_3$  are  $\lambda_{1,2} = \frac{-2b \pm \sqrt{(b^2 - 4ac)}}{2}$  and  $\lambda_3 = -\alpha_3$ , where  $b = k_1x_2 + k_1x_1 + \alpha_1 + \alpha_2$ ,  $a = 1$ ,  $c = k_1x_2\alpha_2 + k_1x_1\alpha_1 + \alpha_1\alpha_2$ . Since  $a, b, c > 0$ ,  $\sqrt{b^2 - 4ac}$  is either complex or  $< b$ . Hence, the real parts of  $\lambda_{1,2}$  are always negative.

$$J_3 = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_3} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_3} \end{pmatrix} = \begin{pmatrix} -k_1x_2 - \alpha_1 & -k_1x_1 & 0 \\ -k_1x_2 & -k_1x_2 - \alpha_2 & 0 \\ k_1x_2 & k_1x_1 & -\alpha_3 \end{pmatrix} \quad (2.6)$$

It is apparent from the above example that we need a more sophisticated method for reaction networks whose dynamics follows mass action law.

## 2.5 Arbitrarily large dynamic systems and mass action kinetics

From the example shown in 2.5 it may appear that the reaction systems following mass action kinetic models have rather fragile dynamics when

compared to the ones which follow Michaelis Menten kinetic models. However, a series of mathematical theories published since 1972 have shown the opposite. Most of these theories are based on the works of [Feinberg \(1972\)](#), [Horn and Jackson \(1972\)](#) and [Volpert and Hudjaev \(1985\)](#). Though [Volpert and Hudjaev \(1985\)](#)'s theory came out more than a decade after that of [Feinberg \(1972\)](#) and [Horn and Jackson \(1972\)](#) we shall discuss it first to keep consistency with our earlier study in this chapter. First we shall describe the conceptual ideas of [Volpert and Hudjaev \(1985\)](#)'s theorem without going into mathematical details (2).

**Theorem 2.** *The solutions of an acyclic reaction system which follows mass action law have the following properties:*

- *The solutions are defined for  $t > 0$ .*
- *There does not exist a nonnegative periodic solution.*
- *The solutions converge to an equilibrium as  $t \rightarrow \infty$ , i.e. the solutions are asymptotically stable.*
- *The equilibrium points or the steady states of such systems are always non-negative ( $\geq 0$ ).*

[Volpert and Hudjaev \(1985\)](#)'s theory is a powerful result in chemical reaction network theory. [Siegel and Johnston \(2003\)](#) has further extended these results to show that a network of reversible reactions also has the same properties as an acyclic chemical reaction network. A chemical reaction system made of reversible reactions is often called 'reversible

system' in the literature regarding chemical reaction network theory. The original work of [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) prior to [Volpert and Hudjaev \(1985\)](#) discussed the dynamic properties of a more generalised class of chemical reaction network, called 'weakly reversible networks' as termed by them. We shall further generalise these results for a more practical class of networks which we call 'partially reversible networks'. Partially reversible networks consist of both reversible and irreversible reactions. We shall discuss some of the concepts introduced by [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) and their main results before going into our extension of their theory.

[Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) developed a consistent mathematical theory to predict the dynamic behaviour of a chemical network from its structural properties rather than having to model their dynamics explicitly. Though their theories are widely used in chemical engineering their application is rarely found in the systems biology literature. In their theory, [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) discussed the dynamic properties of networks of chemical reactions which follow mass action law. The commonalities of our problem and the problem pursued by [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) are apparent. However, most of the theories developed by [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) are applicable to finite and rather small networks of reactions. In our case, we want to determine the dynamic behaviour of arbitrarily large networks which follow the mass action kinetic models. We shall use many ideas and theorems deduced by [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) to develop our theory. The notations that we shall use to describe [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#)'s theory



and our extension are described as follows.

We shall describe the notations used in the following analysis in the light of the simple example discussed before(2.5). Let us consider the chemical reaction system shown in equation 2.5. The set of vertex  $V_3 = \{v_1, v_2, v_3\}$  represents the set of biochemical species involved in the network. The vector of concentrations of the species set  $\{v_1, v_2, v_3\}$  can be given in a vector space  $\mathcal{R}^{V_3}$  whose basis is spanned by the species  $v_1, v_2, v_3$ . Since the concentrations of the biochemical species are always nonnegative the representation of the concentration vector can be constrained into the vector space  $\overline{\mathcal{P}}^{V_3}$ .  $\overline{\mathcal{P}}^{V_3}$  is called the species space of the network shown in figure 2.4 and described by the reaction shown in equation 2.5. The complexes of reaction 2.5 are the left and right hand side of the equation 2.5. Hence, in this case the complexes are  $y_1 = v_1 + v_2$  and  $y_2 = v_3$ . In the complex space, the reaction of equation 2.5 takes the form of equation 2.7.



The set of complexes involved in reaction 2.7 is given by  $Y_3 = \{y_1, y_2\}$  and the vectors representing the states of the complexes are given in the space  $\mathcal{R}^{Y_3}$ . On the other hand, each complex can be represented as a linear combination of all the species, i.e.  $y_i = \sum_{j=1}^3 \gamma_{ij} v_j : i = \{1, 2\}$ , where  $\gamma_{ij}$  are nonnegative integers which are stoichiometric coefficients of the chemical reaction 2.5. In this case  $\gamma_{11} = 1, \gamma_{12} = 1, \gamma_{13} = 0, \gamma_{21} =$

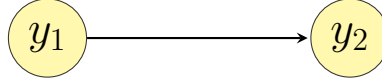


Figure 2.6: Representation of the chemical reaction network shown in figure 2.4 in complex space.

$0, \gamma_{22} = 0, \gamma_{23} = 1$ . Hence,  $y_{1,2} \in \mathcal{N}^{V_3}$  where  $\mathcal{N}$  is the set of nonnegative integers. Each reaction network in species space forms a network of reaction in the complex space too. For example, the chemical reaction network shown in figure 2.4 takes the form of figure 2.5 in the complex space. The reactions rate constants in the complex space remains the same as the rate constants in the species space. Since the reaction  $v_1 + v_2 \rightarrow v_3$  has a rate constant  $k_1$ , the equivalent reaction in the complex space has the same rate constant, i.e. the reaction  $y_1 \rightarrow y_2$  too has a rate constant  $k_1$ . [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) have shown that the dynamics of a chemical network (or biochemical network in this case) in the species space has exactly the same properties as that in the complex space. But before going into any more detail we shall formally define some of the basic concepts presented by [Feinberg \(1972\)](#); [Horn and Jackson \(1972\)](#) for generic networks of chemical reactions. Some of these definitions are also discussed in [Gunawardena \(2003\)](#).

**Definition 4.** A complex  $y$  is a linear combination of species  $y = \sum_{i=1}^N \gamma_i v_i$  of a biochemical network consisting of  $N$  numbers of species. Here  $\gamma_i \geq 0 : i = 1 \dots N$  are stoichiometric coefficients. The projection of  $y$  on the  $i^{th}$  species  $v_i$  is denoted by  $y_{v_i} = \gamma_i$ . Hence,  $y_{v_i}$  is the stoichiometric coefficient of the species  $v_i$  in the complex  $y$ .

**Definition 5.** A biochemical reaction network is a quadruple  $S = (V, Y, R, \kappa)$  where  $V$  is a finite set of species;  $Y$  is a finite set of multisets of species, called complexes;  $R$  is a relation on  $Y$ , denoted  $y \rightarrow \dot{y}$  for

## 2.5 Arbitrarily large dynamic systems and mass action kinetics

---

$y, \dot{y} \in Y$ , which represents a reaction converting  $y$  to  $\dot{y}$ ; and  $\kappa : R \rightarrow \mathcal{P}$  associates a positive rate constant to each reaction.

The dynamics of a biochemical reaction network  $S$  can be given as follows

$$\frac{dc}{dt} = \sum_{y \rightarrow \dot{y} \in R} k_{y \rightarrow \dot{y}} f(y) (\dot{y} - y) \quad (2.8)$$

Here  $c \in \mathcal{R}^V$ ,  $f(y) = \prod_{v \in V} c_v^{y_v}$ ,  $k_{y \rightarrow \dot{y}}$  is the rate constant for the reaction  $y \rightarrow \dot{y}$ .

Let  $\mathcal{R}^Y$  denotes complex space. It has a natural basis defined by the characteristic functions of the singleton subsets  $\{y\}$ , where  $y \in Y$ . If  $U \subset Y$  let  $\omega_U : Y \rightarrow \{0, 1\}$  denote its characteristic function:

$$\begin{aligned} \omega_U(y) &= 1, y \in U \\ &= 0, otherwise \end{aligned} \quad (2.9)$$

For a singleton set,  $\{y\}$  where  $y \in Y$ , we shall use  $\omega_y$  in place of  $\omega_{\{y\}}$ . Complex space then has the standard basis  $\omega_y | y \in Y$ . Let  $\zeta : \mathcal{R}^Y \rightarrow \mathcal{R}^V$  be the linear map defined by  $\zeta(\omega_y) = y$ . Let  $\Psi : \mathcal{R}^V \rightarrow \mathcal{R}^Y$  defined by  $\Psi(y) = f(y) = \prod_{v \in V} c_v^{y_v}$ . These two maps enable us to move back and forth between species space and complex space. It remains only to encode the dynamics on complex space. Let  $x \in \mathcal{R}^Y$  and suppose that  $x$  has components  $x_y$  with respect to the standard basis  $\omega_y$ , so that  $x = \sum_{y \in Y} x_y \omega_y$ . For a given kinetics define  $\kappa : R \rightarrow \mathcal{P}$  define the map

$A_\kappa : \mathcal{R}^Y \rightarrow \mathcal{R}^Y$  by

$$\sum_{y \rightarrow y_1 \in \mathcal{R}} (k_{y \rightarrow y_1}) x_y (\omega_{y_1} - \omega_y) \quad (2.10)$$

Note that  $A_\kappa(x)$  is evidently linear in  $x$ . It can now be seen that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{R}^Y & \xleftarrow{A_\kappa} & \mathcal{R}^Y \\ Y \downarrow & & \uparrow \Psi \\ \mathcal{R}^V & \xleftarrow{f} & \mathcal{R}^V \end{array} \quad (2.11)$$

In other words,  $f = \zeta A_\kappa \Psi$ .

**Definition 6.** A fixed point of a chemical reaction network is a state  $c \in \mathcal{P}^V$  for which  $\frac{dc}{dt} = 0$ .

A fixed point is where  $f = 0$ . This may come about only in certain ways along the composition  $\zeta A_\kappa \Psi = 0$ . If  $c \in \mathcal{P}^V$  then  $\Psi(c) \in \mathcal{P}^Y$ , so  $\Psi(c) \neq 0$ . The next possibility is that  $A_\kappa \Psi(c) = 0$ , i.e.  $\Psi(c)$  is a kernel (not to be confused with the biological kernel discussed in the introduction) of  $A_\kappa$ . According to [Feinberg \(1972\)](#) such phenomena occur when a chemical reaction is weakly reversible and has zero deficiency. The definition of weak reversibility and deficiency of a chemical network and their significance is discussed in detail later in this section. However, the third possibility of occurring a fixed point in a chemical reaction network

is when  $A_\kappa \Psi \neq 0$  but  $\zeta A_\kappa \Psi = 0$ . In that case the solutions of the system are the kernels of the transformation matrix  $\zeta$  which transforms the states of complexes back into the species space. In this thesis we shall discuss about the situation where a fixed point occurs due to  $A_\kappa \Psi = 0$  because as mentioned by [Feinberg \(1972\)](#), this situation occurs for a large class of chemical networks including the type of network we intend to analyse. We shall discuss about this later in the section, but, before going into any more detail let us provide few more definitions regarding stoichiometric subspace, linkage classes and deficiency of a biochemical or chemical reaction network.

**Definition 7.** *The stoichiometric subspace of a reaction network is the vector subspace of  $\mathcal{R}^V$  defined by  $S = \text{span}\{\dot{y} - y | y \rightarrow \dot{y}\}$ . Since  $\frac{dc}{dt}$  is a linear combination of  $\dot{y} - y$ ,  $\frac{dc}{dt} \in S$ .*

**Definition 8.** *The linkage classes of a chemical reaction network are the connected components of the standard reaction diagram formed by assuming complexes as nodes and reactions as edges.*

**Definition 9.** *The deficiency of a chemical reaction network  $\delta$  is defined as  $\delta = m - l - s$ , where  $m$  is the number of complexes,  $l$  is the number of linkage classes, and  $s$  is the dimension of stoichiometric subspace.*

**Definition 10.** *A reaction network  $S$  is weakly reversible if  $\forall y, y_1 \in L, y \Rightarrow y_1$  and  $y_1 \Rightarrow y$  in  $S$ , where  $L$  is a linkage class of  $S$ .*

Given the above definitions the properties of fixed points or steady states of a chemical reaction network can be given by Feinberg's zero deficiency theorem ([Siegel and Chen \(1995\)](#), [Feinberg \(1972\)](#)) as stated below.

**Theorem 3.** *For any reaction with deficiency zero the following statements hold true [Feinberg \(1972\)](#).*

- 1. If the network is not weakly reversible then for arbitrary kinetics (not necessarily mass action law) the differential equations of the corresponding reaction system cannot have either positive steady states or periodic solutions along which the concentration of all the species remains positive.*
- 2. If the network is weakly reversible, then for mass action kinetics (but regardless of the values of the kinetic parameters until and unless they are positive) the set of differential equations of the corresponding reaction system have the following properties: there exist within each stoichiometric compatibility class precisely one positive steady state which is asymptotically stable, and there exist no periodic solution along which all species concentrations are positive.*

Notice that [Feinberg \(1972\)](#)'s theorem is applicable to weakly reversible systems only. From the definition of weak reversibility (definition 10) it is clear that a chemical network can be weakly reversible if and only if each of its complexes is connected to a feedback loop. However for the time being we are mainly interested in acyclic network structures. Hence in the next few sections we shall discuss the applicability of Feinberg's zero deficiency theorem ([Feinberg \(1972\)](#)) on arbitrarily large acyclic biochemical networks which follow the laws

of mass action kinetics. We shall first investigate whether such networks fulfil the two main requirements of zero deficiency theory, i.e. whether the deficiency of these networks are zero and they are weakly reversible or not. There are existing theories, e.g. that of Siegel and Chen (1995) which suggests that a certain class of acyclic reaction network has zero deficiency. Following Siegel and Chen (1995) we discuss those acyclic networks which have zero deficiency. Siegel and Chen (1995) has shown that the deficiency of an biochemical network is zero if its S-C-L (species-complex-linkage) graph is acyclic.

**Definition 11.** *The support of a complex (reaction vector)  $y = \sum_{i=1}^N \gamma_i v_i$  is a set of species  $\text{supp}(y) = V_i = \{v_i | \gamma_i \neq 0\}$ . In other words, the support of a complex (reaction vector) is the set of the species it is made of.*

**Proposition 6.** *The support of any linear combination of the reaction vectors of a linkage class must have more than 2 elements.*

**Proof:** The supports of the complexes of a linkage class are mutually exclusive. Hence, there must be at least two complexes present in the linear combination of the reactions vectors. Since, each complex must consist of at least one species the resultant vector must have a support of cardinality  $> 2$ .

**Definition 12.** *The S-C-L graph of a reaction network is a bipartite graph whose vertex set is partitioned into the species set  $V$  of the network and the set of its linkage classes  $\mathcal{L} = L_i : i = 1 \dots q$  assuming that the reaction network has  $q$  numbers of connected components and whose edges are draw as follows. For each complex  $y \in Y$  in which a species*

*appears draw an edge between the species and the linkage class which contains the complex  $y$ . In addition label each edge by writing the complex for which it was drawn.*

**Proposition 7.** *According to Siegel and Chen (1995) an S-C-L graph is acyclic if and only if the none of the following is true.*

1. *Some species appear more than once within the same linkage class*
2. *There exists at least one pair of species sets corresponding to different linkage classes such that their interaction contains at least two distinct species.*
3. *There exists at least one group of species set  $V^i : i = 1 \dots p$  corresponding to a set of linkage classes  $L_i : i = 1 \dots p$  such that  $V^1 \cap V^2 = v_{12}, V^2 \cap V^3 = v_{13} \dots V^p \cap V^1 = v_{p1}$  for some distinct species  $v_{i,(i+1)}, 1 < i < p - 1, v_{p1}$ .*

**Lemma 2.** *If the S-C-L graph of a biochemical reaction network is acyclic then it has a deficiency zero.*

**Proof:** If the S-C-L graph is acyclic then each linkage class is acyclic with no repeating species. The minimum spanning tree of each linkage class has exactly  $n_i - 1$  edges where  $n_i = |Y_i|$  and  $Y_i$  is the set of complexes involved in the reactions of linkage class  $L_i$ . Each edge  $y \rightarrow y_1; y, y_1 \in Y_i$  corresponds to a stoichiometric vector  $y_1 - y$ . Since no complex in  $Y_i$  contains a common species  $y_1 - y$  are linearly independent vectors  $\forall y, y_1 \in Y$ . Hence, there are exactly  $n - 1$  linearly



independent stoichiometric vectors in every linkage class of the reaction network. For each linkage class  $L_i$ ,  $\delta_i = n_i - l_i - s_i$ , where  $n_i$  is the number of complexes present in the linkage class  $L_i$ ,  $l_i$  is the number of linkage classes, in this case  $l_i = 1$ ,  $s_i = n_i - 1$  is the dimension of the stoichiometric subspace. Hence,  $\delta_i = n_i - 1 - (n_i - 1) = 0$ .

The deficiency of a network is given by  $\delta = \sum_{i=1}^p m_i - l - s = \sum_{i=1}^p (m_i - 1) - s = \sum_{i=1}^p s_i - s$ . Here  $s_i$  is the dimension of the stoichiometric subspace and  $m_i$  is the number of complexes of the  $i^{th}$  linkage class, and  $s$  is the dimension of the stoichiometric subspace of the entire network. Since  $\delta \geq 0$ ,  $\sum_{i=1}^p s_i \geq s$ . The stoichiometric subspace of the entire network is smaller than the total stoichiometric subspaces of the linkage classes if and only if the stoichiometric subspaces of the linkage classes are linearly dependent.

Let us consider that there exist a stoichiometric vector  $s\tilde{v}_p$  in the  $p^{th}$  linkage class which is a linear combination of the stoichiometric vectors of the other linkage classes.

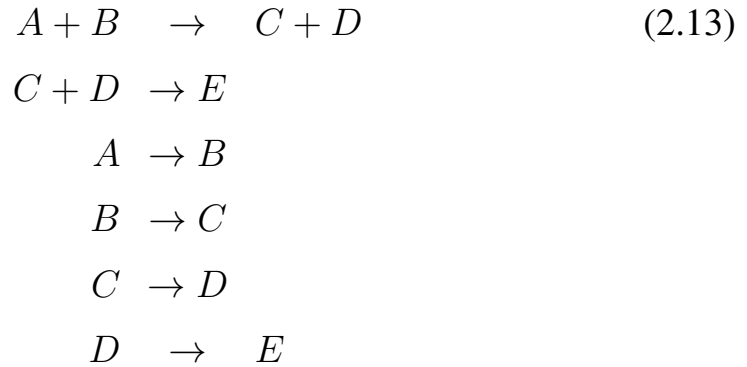
$$\begin{aligned} s\tilde{v}_p &= \sum_{i=1}^p \sum_{j=1}^{m_i-1} a_{ij} \tilde{y}_{ij} \\ &= \sum_{i=1}^p d_i \tilde{s}v_i \end{aligned} \tag{2.12}$$

In equation 2.13  $\tilde{y}_{ij}$  is  $j^{th}$  stoichiometric vector of the  $i^{th}$  linkage class,  $a_{ij}$  are linear coefficients,  $\tilde{s}v_i$  are the linear combination of the reaction vectors of the  $i^{th}$  linkage class.

Let  $L_i$  be the linkage class corresponding to  $s\tilde{v}_i$ ,  $V_i$  be its species

set. Let  $G$  be the graph with vertices  $L_i$  and edge  $e_{ij}$  links  $L_i$  with  $L_j$  if  $|V_i \cap V_j| \geq 1, i \neq j$ . Since by proposition 7  $|V_i \cap V_j| \leq 1$ . Since  $\text{supp}(v_i) \geq 2, \forall i \exists i_1 \neq i_2 \neq i$  such that  $|V_i \cap V_{i_1}| = 1, |V_i \cap V_{i_2}| = 1$ . Hence,  $G$  is cyclic which violates condition 3 of proposition 2. Since the reaction vectors of the linkage classes are linearly independent the deficiency of such network is always zero.

The above theory considers only those networks whose S-C-L graph is acyclic. Though such networks cover a large classes of acyclic reaction network it is however not possible to show that the zero deficiency phenomena is true for all acyclic networks. In fact it can be easily shown that there exist certain classes of acyclic networks whose deficiency are grater than zero. For example consider the following reaction system.



It can easily be shown that the reaction vector  $C + D - A - B$  is linearly dependent on the vectors  $B - A, C - B, D - C$  and hence the system has a deficiency greater than zero. However such networks fall into a rather small and degenerate class of acyclic graphs.

Though the networks whose S-C-L graphs are acyclic, have a zero

deficiency, they are by no means either reversible or weakly reversible. However, if all the reactions of the above network were reversible then the system would have been both reversible and zero deficiency and hence all of its stoichiometric compatibility class would have had one asymptotically stable equilibrium (Feinberg (1972); Horn and Jackson (1972)). This is so because assuming reversible reactions only add a negative reaction vector corresponding to each reaction vector and hence the stoichiometric subspace neither grow nor shrink in dimension. But in reality large reaction networks often consist of both reversible and irreversible reactions. In that case clearly the stoichiometric subspace does not change in dimension, however the system does not remain reversible or even weakly reversible. In the next theorem we show that such reaction networks are weakly reversible under certain assumptions and hence always converge towards asymptotically stable equilibria.

**Theorem 4.** *A system of partially reversible reactions are weakly reversible and have zero deficiency under the following assumption. If the sources (complexes with indegree zero) of a chemical reaction network have a constant source and the sinks (complexes with outdegree zero) are degradable then the corresponding network have zero deficiency and weak reversibility.*

The assumption of theorem 4 is biologically more realistic than assuming total reversibility in reaction network (Gunawardena (2003)). Under the above assumption all the complexes lies at the bottom of the reaction hierarchy are constantly degradable where as there are constant supplies of the reactants which lie on the top layer. Let us assume that  $Y_r$  and  $Y_t$  are the sets of root (source) and terminal (sink) complexes in

## 2.5 Arbitrarily large dynamic systems and mass action kinetics

---

a reaction hierarchy. Under the assumption of the theorem 4  $\forall y_t \in Y_t$  the following reaction exists  $y \rightarrow \emptyset$ . On the other hand  $\forall y_r \in Y_r$  the following reaction exist  $\emptyset \rightarrow y_r$ . If we add the null complex to every linkage class then all the linkage classes will be joined together to form one linkage class. It can be easily seen that the minimum spanning tree of the new linkage class will still have  $m - 1$  edges where  $m$  is the number of complexes in the network including the null complex. Since the supports of all the complexes are mutually exclusive the network will have a deficiency zero.

Since each linkage classes are connected components the terminal nodes can be reached from at least one root node. Again, since there exist at least one path from a root node to a terminal node via any other node any node can be reached from any other node via the null complex. This is due to the fact that the null complex connects the terminal nodes of a reaction network with its source nodes. Hence such a reaction network retains weak reversibility.

Since partially reversible networks are zero deficiency networks with weak reversibility they always have positive asymptotically stable equilibria.

The summary of the above theories and findings is as follows.

- Arbitrarily large acyclic dynamic system whose dynamics follow Michaelis Menten kinetics will always converge to a stable equilib-

rium. But whether the network has a global attractor or not is not clear. Since such systems are nonlinear, there is a possibility that they might have more than one stable state but no unstable or oscillatory state. This is a very different phenomenon than bistability or multistability since there exists at least one unstable equilibrium in bistable or multistable systems. It is the unstable equilibria which help the system bifurcate between different steady states. The lack of such unstable steady states make the system non-bifurcative.

- If there are feedback loops present in such a network then the dynamics of the network is determined solely by the feedback loops. The feedback loops in a Michalis Menten type reaction system are rather sensitive to parameter values and can bifurcate between different dynamic modes. Their bifurcation, however, depends mainly on their type, and their own parameter values and can be called local bifurcation.
- Mass action reaction systems are much more stable compared to Michalis Menten kinetic models. It has been shown that all acyclic mass action systems have at least one asymptotically stable equilibrium, so does most reversible, partially reversible and weakly reversible systems. Since weak reversibility of a mass action network comes from feedback interactions only it might appear that mass action networks are stable for even a large class of cyclic networks. This, however, is not true when it comes to biological systems. The primary condition of stability of weakly reversible mass action system is that the system has to be a zero deficiency system. Zero deficiency systems are irreducible. But biochemical networks are reducible systems (Clements et al. (2009); Liu and

Ochman (2007) and Adami (2007)). In case of reducible systems, Feinberg (1972); Horn and Jackson (1972) and Volpert and Hudjaev (1985) have shown that even in mass action kinetics bifurcation can only arise from cyclic structures. Hence in case of reducible systems the bifurcation is originated and determined by the cyclic structures only.

- From the analysis we see that no matter what kind of kinetic laws a network follows it is always the feedback mechanisms which are responsible for its bifurcation. Without feedback loops such networks will not be bifurcative and will always converge towards asymptotically stable equilibria.

## 2.6 Interpretation of the dynamic properties of biochemical networks

Before going into the biological interpretations of the above theorems we describe the type of dynamics exhibited by acyclic Michaelis Menten and mass action system. It has been shown that they can only converge to stable or asymptotically stable steady states but not to unstable or oscillatory states. However whether they always converge to the same steady state or not is not yet clear. In fact, Horn and Jackson (1972) and Feinberg (1972) at first thought that they proved just that. Hence, in their original paper the theorem states that the above type of networks converge towards global attractors. Later Volpert and Hudjaev (1985)

## 2.6 Interpretation of the dynamic properties of biochemical networks

---

proved that it is only true for a particular class of networks but a general proof does not exist. Hence the global attractor theorem remains a conjecture rather than a theorem. The difference between global and local attractors are as follows:

Imagine a concave surface. If a small sphere is left anywhere in the concave surface it will always roll down to the centre of the concave. In this case the centre of the concave can be thought of as a global attractor. Now imagine a plane horizontal surface. If a sphere is left on the plane it will remain there until certain force is applied to move it to another place. If such a force is applied then the sphere will move to a certain distance and then finally come to rest at a different point. Hence both the initial and the final position of the sphere are local attractors. In fact every possible point on such a plane is a potential local attractor. Which attractor will be chosen by the sphere depends on the applied force. By this analogy the dynamics of an acyclic or reversible or weakly reversible biochemical network is similar to that of a sphere on a plane which is made by joining some concave surfaces together so that there is no space between their peripheral boundaries. It is assumed by **Volpert and Hudjaev (1985)** that solutions do not exist on the peripheral boundaries of these surfaces. The centre of each such concave is a local attractor.

This explanation is true until the Feinberg conjecture is proved. If Feinberg's conjecture is proved then the dynamics of the above type of biochemical networks becomes even simpler. In that case, the dynamics will be similar to that of a sphere on a concave plane. Such systems are termed as monostable systems.

In a cellular environment there are many different kinds of interactions, such as gene regulatory interaction, protein protein interaction, enzymatic reactions, phosphorylation reactions. Some of these interactions have been shown to follow Michaelis Menten kinetics and some of them follow mass action kinetics. The dynamic properties of a reaction network made of both type of reactions can only be determined by developing a general theory of nonparametric chemical kinetics. However, the above analysis suggests beyond doubt that the bifurcations in network, no matter what the reaction type is, are always determined by feedback mechanisms.

## 2.7 The implications of network dynamics on network evolution

How organisms as complex as human beings evolved from a single cellular organism is a fascinating question. One of the fundamental mechanisms of evolutions is driven by random mutations of genetic structures. The changes in the structures of the genes may result in changes in the structures of their product proteins. Such changes in the protein structures may ultimately lead to altered functional behaviour since the function of a protein depends on its structure. A large number of genes and proteins interact with each other in a cellular environment. Changes in the functions of some genes and proteins may affect the dynamic properties of the cellular biochemical interaction network. For



evolution to work, such functional changes are necessary unless they disturb the stability of the entire system. The analysis in this chapter suggests that the global dynamics of a biochemical network is usually shaped by feedback loops present in it. Perturbation in these loops might bring about erratic changes in the bifurcation pattern of a cellular network, i.e. the stability properties of these networks may be altered. Such perturbations might not always be deleterious. But, the possibility of such perturbations to produce deleterious phenotypes is higher in the feedback loops than anywhere else. Hence, we posit that the feedback mechanisms of biochemical networks should evolve at a slower rate compared to the rest of the network. One way to test the above hypothesis is to look for conserved feedback loops in biochemical networks of multiple species. We shall do just that in chapter 4. But before we address the question of applicability of the biochemical reaction network theory on molecular evolution we shall address a corollary question.

The quantity of feedback loops in biochemical networks is important for many reasons. If most of the chemical complexes in such networks are involved in feedback loops then two possibilities might occur. If such a system has deficiency zero then high numbers of feedback mechanisms might in fact ensure asymptotic stability if the network was made fully of mass action type of reactions. However, the essential condition of asymptotic stability is zero deficiency which means the system must be irreducible. Modern systems biology studies suggest that biochemical systems are reducible. Some recent studies regarding reducibility of biochemical systems are as follows, [Adami \(2007\)](#); [Clements et al. \(2009\)](#); [Liu and Ochman \(2007\)](#). Due to reducibility of biochemical systems

presence of too many feedback mechanisms may not help in achieving asymptotic stability and may instead increase the bifurcativeness of the system disregarding the type of reaction present in such a network. Whether highly bifurcative behaviour is desirable in a biochemical network is a different issue but clearly the stability of such systems are sensitive to perturbations. Hence if a biochemical system has too many feedback loops then there is little scope for evolutionary changes to take place without altering the stability of the network which might produce deleterious phenotypes. On the other hand, if these networks have a small number of feedback loops then evolutionary renovations may take place more freely in the acyclic parts of these network without disturbing the stability of the network. Hence as a corollary question of the analysis pursued in this chapter, we shall first investigate the percentage of genes and proteins involved in feedback mechanisms in a typical biochemical network. If we find that a rather small percentage of genes and proteins form feedback structures than our assumption regarding the effects of dynamic properties on the evolution of a biochemical network will make sense.

## Chapter 3

# Structural properties of biochemical networks

In this chapter we shall determine the amount of feedback loop in the biochemical networks of *S. cerevisiae*. Biochemical networks are made of many different kinds of interactions, such as gene regulatory interactions, protein protein interactions, metabolic interactions etc. The most detailed picture of a biochemical network can be obtained by taking all these interactions into account. However, doing so will make any network analysis extremely complicated. Most of the studies on the structural properties of biochemical networks take into account only one kind of interaction for the sake of simplicity. For example, **Yu and Gerstein (2006)** analysed the gene regulatory network of *S. cerevisiae* and concluded that these networks have hierarchical structures. Such hierarchy arises due to highly acyclic nature of these networks. The studies

---

by Yu and Gerstein (2006) include only interactions between transcription factors and their target DNA binding sites (TF-DNA), and ignore other biochemical interactions. On the other hand there are several other studies which focussed only on the protein interaction networks, e.g. Schwikowski et al. (2000). So far there has been relatively few efforts on analysis of the integrative networks of gene regulations and protein interactions. One such effort is by Yeager-Lotem et al. (2004) who built a collective model of yeast biochemical network by integrating TF-DNA interactions with protein interactions. Though Yeager-Lotem et al. (2004)'s method is used in some other studies regarding structural analysis of biochemical networks it has some drawbacks. The main problem with Yeager-Lotem et al. (2004)'s method is that it does not consider the fact that the structures of biochemical networks change between different phases in the yeast life cycle. The first comprehensive analysis of the changes in the structure of gene regulatory networks of *S. cerevisiae* during different phases of its life cycle was pursued by Luscombe et al. (2004b). Luscombe et al. (2004b) reconstructed the gene regulatory networks of yeast for different phases of sporulation and cell cycle and found that the reconstructed networks had very different structural properties. However, Luscombe et al. (2004b) considered only TF-DNA interactions in their study and ignored other kinds of interactions such as protein protein interactions. The most accurate information about the structural properties of a biochemical network can be gained by analysing the integrated networks of gene regulations and protein interactions specific to different cellular phases of yeast life cycle.

Before going into details we shall discuss the outline of our method

---

of structural analysis of biochemical networks. A workflow diagram for our structural analysis is shown in figure 3.1.

The main steps of the structural analysis are as follows.

### **Inferring cellular phase specific biochemical network**

We used the analysis of [Luscombe et al. \(2004b\)](#) as the backbone of our analysis. [Luscombe et al. \(2004b\)](#) divided the life cycle of budding yeast in many different cellular phases such as metabolic phase; early 1, 2, early mid, middle and mid late phases of sporulation and early, late G1; G2, M and S phases of cell cycle and inferred the subset of transcription regulations that take place during each of these cellular phases. We used this data to infer protein interaction networks specific to each of the cellular phases mentioned by [Luscombe et al. \(2004b\)](#). The most obvious way to infer protein interactions from gene regulation data is to assume that the protein products of highly expressed genes are also present in high concentration in the cytoplasm. Recently [Bharhadwaj and Lu \(2005\)](#); [Rogers et al. \(2008\)](#) find that the protein concentrations are rather uncorrelated to the concentrations of their corresponding mRNAs. Hence, we are left with little choice but to consider gene ontology annotations as features for the purpose of assigning protein protein interactions to specific cellular phases. Since gene ontology annotations provide information about the biological processes, the cellular compartments and the molecular functions of proteomic and genetic interactions, we choose these annotations as logical choice for the purpose of clus-

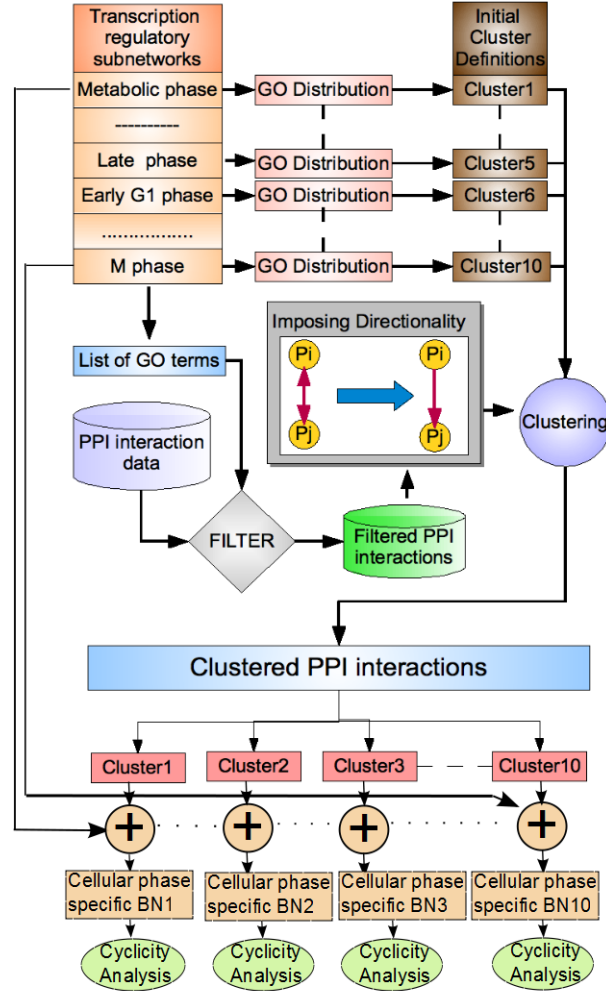


Figure 3.1: Flowchart showing the processing and clustering of protein protein interaction data to form cellular state specific molecular interaction networks. Firstly we made a list of all the GO terms present in [Luscombe et al. \(2004b\)](#)'s dataset. We filtered in all the PPIs which are annotated with the GO terms present in the above mentioned list. Before clustering we applied [Liu et al. \(2009\)](#)'s algorithm to detect directional interactions (such as phosphorylation, dephosphorylation, acetylation etc.) and the relevant directions of signal flow, i.e. which protein phosphorylates whom etc. Once the directional interactions are identified and directions of interactions are determined we clustered the PPI data. The GO distribution of each cellular state specific subnetwork is considered as prior observations and used as initial cluster definition. The GO annotation of each PPI is then used as feature for the clustering algorithm and the filtered set of PPIs are clustered in different cellular phases. Finally the clustered PPI subnetworks are merged with their corresponding GRNs to make cellular phase specific integrated biochemical network. The fractions of proteins and DNAs of each subnetwork which take part in cyclic structures are then calculated.

---

tering. We cluster binary protein interaction data taken from the curated interaction database BIND (Bader et al. (2001)) based on their gene ontology(GO) annotations. For clustering, we consider the GO distributions of the cellular phase specific GRNs Luscombe et al. (2004b) as prior observations and select the protein interactions which have similar GO distributions. The detailed methodology is described in section 3.1 and 3.2. Finally, following Yeager-Lotem et al. (2004) we integrated the inferred PINs to their corresponding GRNs to build 10 different biochemical networks which corresponds to 10 cellular phases in yeast life cycles (Luscombe et al. (2004b) reported 12 cellular phases but we merged three of them together for convenience of statistical inference) We analysed the amount of feedback mechanisms in these biochemical networks.

### **Inferring direction of signal flow in cellular phase specific biochemical networks**

Our goal in this chapter is to determine the amount of feedback loops present in the cellular phase specific biochemical network. The integrated cellular phase specific biochemical networks contain both directed (TF-DNA interactions) and undirected (protein-protein interaction) edges. To infer the relevant structural properties we need to know the direction of signal flow in the protein interaction networks. Thus, we inferred the direction of signal flow in the cellular phase specific protein interaction networks from their domain interaction information as suggested by Liu et al. (2009). This yielded a complete picture of signal flow

---

through protein interaction networks and how they affect gene regulation and vice versa in each cellular state of yeast life cycle. The resulting networks are fully directed graphs which we finally used to determine the amount of feedback loops their architecture. The methodology of inferring the signal flow direction in a protein interaction network is shown in section [3.3](#).

### **Determining the amount of feedback loops in cellular phase specific biochemical networks**

To quantify the relative abundance of cyclic and acyclic motifs we first enlist all the feedback mechanisms present in each of the inferred networks. Then we calculate the ratio between the number of nodes(i.e. genes or proteins) involved in feedback mechanisms and the total number of nodes. Multiplying these numbers by 100 gives us the percentage of genes and proteins which forms feedback loops. These percentages represent the amount of cyclicity in a biochemical network.

### **Determining the effect of data specific and sampling noise on our calculation**

The percentages calculated above are not entirely correct due to presence of noise in the interaction data that we used. The statistical inference methodologies used to recreate cellular phase specific biochemical networks introduced further sampling noise. Hence the inferred networks



are not accurate representation of biochemical networks that take place during different cellular phases of yeast life cycle. There are several efforts to determine the effect of different types of noise on the structural properties of biochemical network. For example [Huang et al. \(2007\)](#) found that the global structural properties such as the diameter of a network do not have significant variations even if 50% of the interaction data accounts for noise. Even before [Huang et al. \(2007\)](#), [Yu and Gerstein \(2006\)](#) have used the same methodology as [Huang et al. \(2007\)](#) to account for noise in their data. Hence we used [Yu and Gerstein \(2006\)](#) and [Huang et al. \(2007\)](#)'s methodology to account for noise in our data. Instead of using the percentages mentioned above we have calculated means and variances of these percentages over an ensemble of networks created by randomly adding, deleting and rewiring 20% of the nodes and edges of the networks as suggested by [Huang et al. \(2007\)](#). This method and the results are discussed in detail in the section [3.4](#).

In the next few sections we shall discuss the above steps in detail.

## 3.1 Network Data Preprocessing

The network data is taken from the curated protein interaction database BIND([Bader et al. \(2001\)](#)). There are over 100000 binary protein interactions in the BIND database. Before clustering, we pre-processed the protein interaction data-set in order to reduce computational cost. In

order to preprocess we first enlisted all the GO annotations of [Luscombe et al. \(2004b\)](#)'s data. We used only the GO terms which are independent of each other. More generic GO terms are not included mainly because almost all the proteins are annotated by the generic terms. We assumed that the leaves of the graph constructed from the interdependencies of GO terms as described in [Ashburner et al. \(2000\)](#) are independent of each other. We find that 37625 interactions are annotated with the same annotations of [Luscombe et al. \(2004b\)](#)'s data. This reduction in number of binary interaction in the filtered set might be questionable since [Luscombe et al. \(2004b\)](#) covers almost all the phases of cellular life cycle. We find that there are almost 15000 independent GO terms in the GO map presented in [Ashburner et al. \(2000\)](#). However, the genes and transcription regulatory interactions mentioned in [Luscombe et al. \(2004b\)](#) are annotated by only 426 of them. There are only 37625 interactions which are annotated by these 426 independent GO terms. Including the more generic GO terms includes more interactions in the filtered set. However, including all the generic GO terms found in the dataset of [Luscombe et al. \(2004b\)](#) filters in all protein protein interaction present in the dataset. This increases the computational complexity in the later stages of our analysis. Hence we kept our dataset limited to the filtered set of 37625 interactions. This set of interactions are then used for clustering.

### 3.2 Clustering protein protein interaction data

Cellular state specific protein interaction networks are constructed based on the Gene Ontology(GO) distribution of the cellular state specific transcription regulatory subnetworks reported by [Luscombe et al. \(2004b\)](#). We clustered binary protein interactions based on their GO distribution using a Bernoulli mixture model based clustering method. Let us assume we have N numbers of protein interactions  $x_i : i = 1..N$ . Since, the set of filtered interactions contains 37625 protein interactions,  $N=37625$ . Each protein interaction  $x_i$  is represented by a binary string of length  $D$ ,  $x_i^j : j = 1..D$ , which represents whether  $j^{th}$  GO term is present in the annotation of the  $i^{th}$  interaction. In our case, we have taken only 426 GO terms into account, hence,  $D=426$ . If  $|\{x_{ij} : x_{ij} = 1\}| = 1$ , i.e. the  $i^{th}$  protein interaction is tagged with only one independent GO term then it is assigned to all the cluster labels where it's GO annotation  $x_{ij}$  is observed apriori. All other interactions are clustered using Bernoulli Mixture Model. To make  $x_i^j$  i.i.ds we considered only the leaf terms of GO networks. If  $x_i^j$  is governed by Bernoulli distribution then  $p(x_i|\mu) = \prod_{j=1}^D \mu_j^{x_i^j} (1 - \mu_j)^{(1-x_i^j)}$ . A finite mixture of these distributions is given by  $p(x_i|\mu_k, \pi) = \sum_{k=1}^K \pi_k p(x_i|\mu_k)$ , where  $p(x_i|\mu_k) = \prod_{j=1}^D \mu_{kj}^{x_i^j} (1 - \mu_{kj})^{(1-x_i^j)}$  and the log likelihood function for this model is given by  $\ln(p(X|\mu, \pi)) = \sum_{i=1}^N \ln\{\sum_{k=1}^K \pi_k p(x_i|\mu_k)\}$ . For the purpose of clustering we introduced a binary latent variable  $z_i = \{z_i^k : k = 1..K\}$  so that  $z_i^k = 1$  if  $x_i$  belongs to the  $k^{th}$  cluster, where  $p(z_i|\pi) = \prod_{k=1}^K \pi_k^{z_{ik}}$  and  $p(x_i|z_i, \mu) = \prod_{k=1}^K p(x_i|\mu_k)^{z_{ik}}$ . Hence, the posterior probability  $p(z_{ik} = 1|x_i) \propto \pi_k p(x_i|\mu_k)$ . We set up a Gibbs sampler with simulated

annealing considering  $p(z_{ik} = 1|x_i) \propto \pi_k \exp(\frac{\ln(p(x_i|\mu_k))}{T})$ , where  $T$  is the temperature parameter. We sampled the posterior while cooling the temperature until we found the optimal clustering allocation for each protein interaction. In this case, the GO distribution of the cellular state specific TF-DNA subnetworks are considered as prior observations.

### 3.3 Directionality in binary protein interaction data

Binary protein protein interaction data are bidirectional. It only tells us which two proteins interact with each other. In order to analyse the amount of cyclic structures or feedback contents in protein interaction networks we need to consider the direction of signal flow through the network since the concept of feedback loop exists only in directed graphs. Introducing the sense of directionality in protein interaction networks makes them directed graphs which might have very different topological characteristics compared to their undirected counterparts. Following [Liu et al. \(2009\)](#) we used domain interaction data to assign signal flow direction to binary protein interaction in our dataset. if  $D_i$  and  $D_j$  represent sets of protein domains present in two interacting proteins  $P_i$  and  $P_j$  then the function

$$F(d_{mn}) = \frac{Pr(d_m \rightarrow d_n) - Pr(d_n \rightarrow d_m)}{Pr(d_m)Pr(d_n)} \quad (3.1)$$

### 3.3 Directionality in binary protein interaction data

---

represents the direction of signal flow when domain  $d_m$  interacts with domain  $d_n$ . Here,  $d_m \in D_i$  and  $d_n \in D_j$  are two protein domains,  $Pr(d_m \rightarrow d_n)$  is the probability of protein interactions in which one protein has  $d_m$  and the other has  $d_n$ ,  $Pr(d_m)$  and  $Pr(d_n)$  are probabilities of  $d_m$  and  $d_n$ . In this case, if  $F(d_{mn}) > 0$  the signal flows from  $d_m$  to  $d_n$  and vice versa. The directionality of protein interactions are determined by

$$\mathcal{P}_{ij} = \frac{\sum_{d_m \in D_i, d_n \in D_j} F(d_{mn})}{N_{ij}} \quad (3.2)$$

where  $N_{ij}$  is the number of domain interactions between  $P_i$  and  $P_j$ . If  $\mathcal{P}_{ij} > 5$  the signal flows from  $P_i$  to  $P_j$  and if  $\mathcal{P}_{ij} < -5$  the signal flows from  $P_j$  to  $P_i$ . The choice of this threshold value is suggested by [Liu et al. \(2009\)](#) since it is found by them that 5 is the minimum threshold which results in high prediction accuracy. The interactions with  $-5 < \mathcal{P}_{ij} < 5$  are left bidirectional. For more details see [Liu et al. \(2009\)](#).

Once the direction of signal flow in the protein interaction network is inferred we integrate each protein interaction network with its corresponding gene regulation network as provided by [Luscombe et al. \(2004b\)](#). Doing so gives us a comprehensive picture of signal flow from the environment through the protein interactions to the gene regulatory networks and vice versa in each cellular phase of yeast life cycle. Though, the above method is supposed to give us accurate pictures of biochemical machinery in yeast life cycle it does not do so in reality. Rather the inferred networks are coarse representation of cellular phase

### 3.4 Relative abundance of cyclic and acyclic structures in cellular state specific biochemical networks

---

specific biochemical networks due to presence of data specific and sampling noise in them. Currently no method exists to quantify the amount of errors in the inferred network. However, some studies such as [Huang et al. \(2007\)](#) discusses the effect of noise on the global structural properties of biochemical networks. In the next section we shall discuss the effect of noise on our results.

### 3.4 Relative abundance of cyclic and acyclic structures in cellular state specific biochemical networks

We quantify feedback mechanisms by calculating the percentage of genes and proteins present in each of the inferred biochemical network. Since our integrated networks contain both data specific and inferential noise we carry out statistical analysis to determine the effect of noise in our estimates. Following [Yu and Gerstein \(2006\)](#), [Huang et al. \(2007\)](#) we constructed 2000 biochemical networks for each cellular states by adding, deleting and rewiring 20% of the vertices and edges and then calculating the mean and variance of the above percentage estimates. The results of our analysis is shown in table [3.1](#). It can be seen from table [3.1](#) that on an average  $10 \pm 5\%$  of genes and proteins take part in feedback mechanisms in a typical cellular phase specific biochemical network, i.e. approximately  $90 \pm 5\%$  genes and proteins form only acyclic architecture. The detailed results of our analysis are shown in table [3.2](#).

### 3.5 Some critical points regarding the above analysis

## Tables

Cellular states	Mean%	Variance	Cellular states	Mean%	Variance
Metabolic phase	4.9%	1.97	Early G1 phase	4.5%	1.88
Early I phase	5.25%	2.24	Late G1 phase	8.4%	3
Early II phase	9.4%	3.47	G2 phase	7.4%	2.64
Early mid phase	7.05%	2.63	M phase	8.25%	2.97
Mid, Mid late, Late phase	12.95%	3.91	S phase	7.4%	2.6

Table 3.1: The percentage of genes and proteins which take part in feedback mechanisms in the cellular state specific biochemical networks of yeast.

## 3.5 Some critical points regarding the above analysis

### 3.5.1 Protein interaction data

In our analysis we used protein interaction data from BIND database ([Bader et al. \(2001\)](#)) which contains both curated and high throughput data. Interaction data resulting from high throughput experiments can be both incomplete and unreliable. [von Mering et al. \(2002\)](#) showed that there is little overlap between some early high throughput protein interaction data sets. This may be due to both incompleteness and high number of false positives present in those datasets. Though the quality of high throughput data have improved recently ( [Gavin et al. \(2006\)](#) ) false positive and false negative interactions in these datasets remain problems. Several attempts have been made to increase the degree of confidence in protein protein interaction data. Some of these approaches use multiple validations assuming that interactions observed in more than

### 3.5 Some critical points regarding the above analysis

---

one experiments are more likely to be true than those which are observed once. Assigning different weights to interactions derived from different types of experiments is also proposed ( e.g. [Suthram et al. \(2006\)](#) ). One of the most reliable yeast protein interaction dataset developed by [Batada et al. \(2006\)](#) also uses multiple validation to reduce the number of false positives. However, [Batada et al. \(2006\)](#) do not address the issues of sampling and biases associated with multiple validations. Multiple validation is in some ways desirable, as it will decrease the false-positive rate. However, it introduces other different kinds of biases. Accepting interactions that are observed at least twice results in a tendency to reject certain classes of interaction data. For example, the set of interactions that will be retained in a multi-validated dataset will tend to be biased towards those that are highly studied. This has a drastic effect on network topology [Hakes et al. \(2008\)](#). To be credible, any biological conclusions drawn from network structure should be robust with respect to the additional validation of interactions. However, [Hakes et al. \(2008\)](#) demonstrated that the removal of interactions results in a radical change of many network properties. The effects of multiple validation will be similar in any dataset until we have unbiased multiple observations of all interactions. Thus, the highest-quality datasets currently available contain interactions which, though reliable, are not necessarily representative of the network as a whole. This means that drawing conclusions about the structure of a network as a whole is problematic even if each of its interactions is rigorously validated.

In our study, we used protein interaction data from BIND database which contains data from many high throughput studies. Since this



### 3.5 Some critical points regarding the above analysis

---

database is a collection of datasets derived from multiple experiments, false positives should be a bigger issue in this case than incompleteness. Large number of false positives may only increase the amount of feedback loops in a network topology. Hence, in the worst case scenario, presence of large number of false positives may have resulted in over estimation of cyclicity in our analysis. Which means the cyclicities of yeast biochemical networks are more likely to be smaller in noise free networks than those which we have estimated. This does not affect the main conclusion of the analysis presented in this chapter. Hence, though the presence of false positives in network data may cause problems in structural analysis of biochemical networks it is unlikely to do so in our analysis.

#### 3.5.2 Assigning directionality in protein interaction data

Another arguably controversial aspect of our analysis is assigning directionality in protein interaction networks. Usually binary protein interactions represent complex formation by physical association of proteins. These type of interactions are bidirectional. Hence, assigning directionality to these interactions does not make much sense. We, used [Liu et al. \(2009\)](#)'s approach to determine directionality in protein interactions. [Liu et al. \(2009\)](#) used the statistics of domain interactions to determine the directionality of signal flow in a protein interaction network. They noticed that the interactions which are directional in nature, such as phosphorylation, dephosphorylation, acetylation etc. have statistically significantly

higher  $P_{ij}$  values (described before in this chapter) compared to the bidirectional interactions. Based on this observation it is then possible to determine which protein interactions in a given database are directional. Since BIND database contains heterogeneous protein interaction data, finding which interactions are bidirectional and which are directional is necessary. However, Liu et al. (2009)'s method is not yet scrutinised by the scientific community and has its own limitations. Hence, the results of using this method can be questioned. In our case, among 37625 interactions only  $\approx 2500$  interactions are found to be directional and assigned a directionality. This consists of only  $< 7\%$  of the entire dataset. Even if some of the inferred directions are wrong they will have too little effect on the network topology to affect the main conclusion of this chapter.

#### 3.5.3 PPI clustering

Clustering protein interactions into functional groups is a common approach used by bioinformaticians in different areas of computational biology Asur et al. (2007). However, clustering them into cellular phase specific subnetworks is not common. We used GO annotation data as feature vectors for clustering. GO annotation data is very noisy. Relying on GO annotation for clustering protein interactions may not be the best approach to find cellular phase specific protein interaction networks. A better approach will be to integrate multiple heterogeneous data types such as genomic context data (gene fusion, gene neighbourhood, and phylogenetic profiles), primary experimental evidence (physical pro-

tein interactions and gene co-expression), manually curated pathway databases, automatic literature mining, protein expression levels, protein phosphorylation levels, GO annotations, domain interaction statistics etc. into a clustering framework. Recently, efforts are being made to integrate heterogeneous data sets to infer functional protein interaction networks [Szklarczyk et al. \(2011\)](#). However, to the best of our knowledge no effort has been made yet to integrate the above types of data to infer cellular phase specific interaction networks. This is mainly due to lack of efforts to investigate at least some of the above factors in the context of different phases of cellular life cycle. Until high quality cellular phase specific heterogeneous datasets become available we have to rely on noisy subsets of features for the purpose of inference. Hence, the networks inferred in this chapter are not accurate representations of cellular phase specific biochemical networks. Instead, they are coarse and noisy representations of the real biochemical events which take place during different cellular phases of yeast life cycle.

### 3.6 Structural properties and evolution of biochemical networks

Our analysis clearly suggests that most parts of the cellular phase specific biochemical networks are made of acyclic interactions patterns. However, it is also apparent from table [3.2](#) that there is considerable number of feedback loops present in these networks. In almost all the networks

### 3.6 Structural properties and evolution of biochemical networks

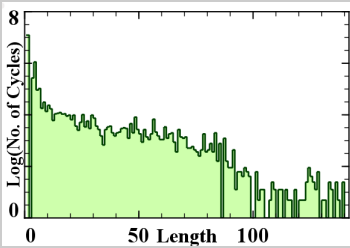
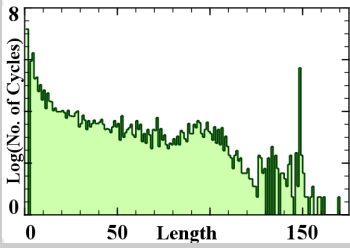
---

there are multiple feedback loops which spans as many as 200-300 genes and proteins. The number and size of the feedback loops detected in these networks might be surprising, especially in view of the fact that a small percentage of genes and proteins are involved in them. This dilemma is perfectly explainable. The maximum number of feedback loops detected in the cellular phase specific biochemical networks is 4582. In theory, this many feedback loops can be generate from only 13 proteins since theoretically the number of feedback loops that can generated from  $N$  number of proteins is  $2^N - 1$ . However we have also detected feedback loops which spans few hundred proteins. But since all these feedback loops are nested and tangled with each other they contain many common genes and proteins between themselves. This explains the relative small percentage of the genes and proteins involved in feedback mechanism. However, it is perfectly possible that such highly connected feedback clusters are results of too many false positives in the analysed data set. We argue that such false positive can only provide us with overestimates of the above calculation, i.e. if the data was in-fact free of false positives we would have found even smaller fraction of genes and proteins to take part in feedback mechanisms. This only suggests that in the real biochemical networks the proportion of acyclic structure are even higher than we have detected. This does not influence our argument in any deleterious way. As discussed in the previous sections, the acyclic parts are rather amenable to evolutionary changes where the feedback loops should be resistant. If most of the biochemical networks are acyclic this only means that most part of a biochemical networks can accommodate evolutionary changes without any immediate deleterious effect. Whereas a rather small proportion of biochemical networks which are

### 3.6 Structural properties and evolution of biochemical networks

involved in feedback mechanism are rather resistant to such changes and should remain conserved for a long period of time.

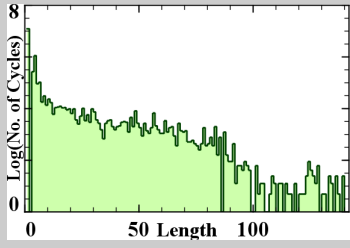
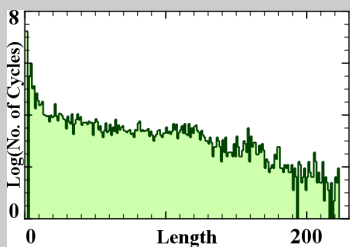
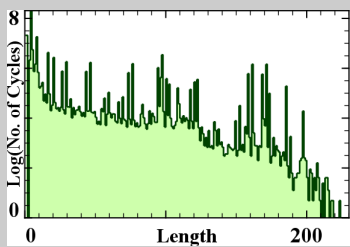
In the next chapter we shall try to find evidence to support the above hypothesis by comparing biochemical networks of yeast species. We shall mainly focus on modules which have feedback mechanisms. We shall compare the conservation of those feedback mechanisms and parts of the acyclic structure surrounding them.

Cellular states	Some of the proteins present in the network	Some of the enriched GO terms	Mean histogram of cycle lengths	% of vertices to form cycle
<b>Metabolic phase</b>	cit1, cit2, cit3, put2, gdh1, sdh4, sdh3, idh1, gtt1, srb2, car2, srb2, alg2, alg8, cdc21, rad9, yah1, swc1, nat4, mig1, med10, spt20, rrp44, ccl1, mdh2, sdh3, aco1, mls1, rav2, rav1, ppa1, vph1, skp1	Glutamate metabolic process, Primary metabolic process, Cellular metabolic process, Regulation of macromolecule metabolic process, Coenzyme catabolic process, Vacuolar acidification, pH reduction		mean: 4.9%, Std. Dev.: 1.97
<b>Early I phase</b>	mus81, rad9, rps3, elg1, pol3, rad16, hof1, mms4, rps3, ski8, far1, sod1, tim18, cdc8, cin8, elg1, cdc13, ste50, hof1, spo11, mcm5, ant1, pol30, pol12, srb2, ant1, ubp2, chs7, msi1, cdc73, psf1, kar3, tof1	RNA metabolic process, DNA metabolic process, DNA replication, DNA repair, Metabolic process, Cell Cycle Process, M Phase of meiotic cell cycle, Meiosis, Regulation of DNA metabolic process, Mitotic recombination, Macromolecule biosynthesis process, Sex determination		mean: 5.25%, Std. Dev.: 2.24

Continued on next page

### 3.6 Structural properties and evolution of biochemical networks

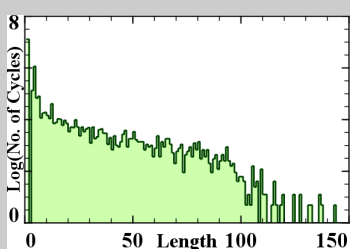
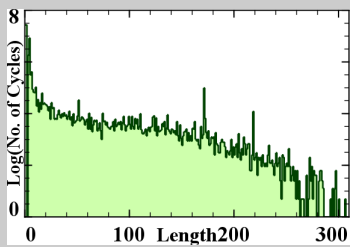
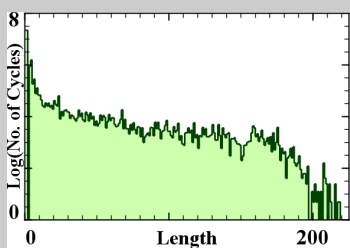
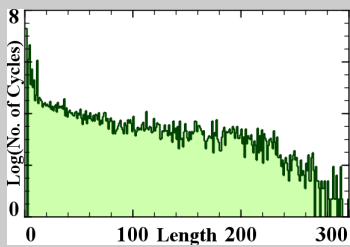
Table 3.2 continued from previous page

Cellular states	Some of the proteins present in the network	Some of the enriched GO terms	Mean histogram of cycle lengths	% of vertices to form cycle
<b>Early II phase</b>	sfb3, cin8, arf2, sar1, pac1, gos1, elp2, sec16, sec4, sec2, kap123, rmd7, sgg1, far1, sar1, lcb4, eri1, syg1, hkr1, zpr1, bem4, gtp1, rdi1, cdc11, ste20, myo5, elm1, cdc3, sfb3, sac6, rvs161, ste2	Establishment of cell polarity, Golgi vesicle transport, Establishment of localization, Establishment / maintenance of cell polarity, Anatomical structure development, Anatomical structure morphogenesis, Cellular structure morphogenesis, Asexual reproduction		mean: 9.4%, Std. Dev.: 3.47
<b>Early mid phase</b>	cdc55, nrg2, ste50, kss1, ste20, rst1, elm1, dfg5, dia1, spt3, ste4, bud5, ste20, pho5, bud8, chs1, rvs161, myo5, las17, dfg5, bem1, bem4, cdc24, chs1, ase1, hof1, gic1, gic2, swi1, msb4	Invasive growth in response to glucose limitation, Asexual reproduction, Regulation of cell size, Pseudohyphal growth, Cellular structure morphogenesis, Anatomical structure development, Cytokinesis, Establishment/ or maintenance of cell polarity		mean: 7.05%, Std. Dev.: 2.63
<b>Middle phase, mid late phase, late phase</b>	cin8, swc1, cdc73, irr1, sif2, sas10, ssn6, rsc1, vps71, rsc2, rox1, elg1, hdf1, tel1, fob1, rtt101, cse4, pnc1, top1, sap30, set1, hst3, rap1, hdf2, fob1, fkh1, set2, set2, yox1, hst1, reg1, nam7, top1, sap30	Chromosome organization and biogenesis, Negative regulation of nucleoside nucleotide and nucleic acid metabolic process, Double-strand break repair, Establishment and/or maintenance of chromatin architecture, DNA packaging, Chromatin assembly or disassembly, Telomere maintenance, etc.		mean: 12.95%, Std. Dev.: 3.91

Continued on next page

### 3.6 Structural properties and evolution of biochemical networks

Table 3.2 continued from previous page

Cellular states	Some of the proteins present in the network	Some of the enriched GO terms	Mean histogram of cycle lengths	% of vertices to form cycle
<b>Early G1 phase</b>	cdc28, swi4, swi5, swi6, hsl1, ste11, fus3, hsl7, swi1, ste7, cdc4, cdc34, cdc45, pho85, pcl1, far1, ste5, pcl5, bub1, skp1, gal11, ctr9, orc1, mcm3, sis2, clb5, bfa1, rmd11, mms4, top3, mcm6, rad24, msil etc.	G1 phase, G1 to G0 transition, G/S transition, Cell cycle, Cell cycle process, Cell cycle phase, DNA integrity checkpoint, Biopolymer biosynthetic process, Cellular component assembly, Cell development etc.		mean: 4.5%, Std. Dev.: 1.88
<b>Late G1 phase</b>	cln1, cln2, cln3, cdc1, cdc15, cdc19, cdc2, cdc20, cdc25, cdc26, cdc28, cdc30, cdc31, cdc33, swi1, swi4, swi5, swi6, swe1, pho11, pho12, pho84, pho85, fus1, fus2, ste20, ste23, ste4, ste6, hsl1, hsl7, gal1, gal11	G1 phase, G/S transition, Cell cycle, Cell cycle phase, Cell division, Metabolic process, Response to stress, DNA repair, DNA integrity checkpoint, DNA replication, DNA dependent DNA replication, Cell budding, Mismatch repair		mean: 8.4%, Std. Dev.: 3
<b>G2 phase</b>	mbp1, swi4, swi6, abp1, actin, bap60, bap55, bat2, bbp1, clb2, bck1, bdf2, swe1, cbp6, dbf2, dbf4, dcc1, ddef1, ddef2, sh3kbp1, act1, cg7846, mad212, cln1, ymr31, pre8, shu1, mrp4, npr2 etc.	G2 phase, G2/M transition DNA damage checkpoint, G2/M transition size control check point, Cell cycle, Cell cycle process, Cell cycle phase, Cell division, DNA repair, Spindle organization and biogenesis, Cytokinesis, Chromosome segregation		mean: 7.4%, Std. Dev.: 2.64
<b>M phase</b>	cin8, rim4, ime1, nap1, ask1, ubc9, ccl1, mms22, mms4, sae3, ctf4, mcm2, cdc16, pms2, rad24, med9, ypk1, ssl1, rev7, mob2, spc25, tom1, nup53, spo12, ctf8, cdc5, apc4, ctf18, cse1, mlh1, zip3, cdc27, mer1, msh2, cse2, rad51 etc.	M phase, M phase of mitotic cell cycle, Cell cycle, Cell cycle process, Cell cycle phase, Cell division, Mitotic metaphase anaphase transition, transcription, Regulation of transcription, Cyclin catabolic process, Mitotic cell cycle		mean: 8.25%, Std. Dev.: 2.97

Continued on next page

### 3.6 Structural properties and evolution of biochemical networks

Table 3.2 – concluded from previous page

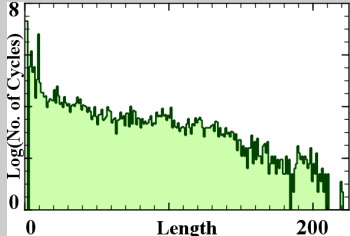
Cellular states	Some of the proteins present in the network	Some of the enriched GO terms	Mean histogram of cycle lengths	% of vertices to form cycle
<b>S phase</b>	rsc6, bap60, bem2, sit4, sap190, sap155, tap42, sap185, dbf2, dbf20, bik1, kem1, ski8, ski3, lsm3, lsm5, lsm6, lsm7, pat1, psu1, dbf4, kar3, cdc73, rrp6, ctf8, chs1, pdc1, pph3, rrd1, sit4, adr1, ppt1, mac1, sti1, rppl1, hsp82, ssa1.	G/S transition, S phase, S phase of mitotic cell cycle, Cell cycle, Cell cycle process, mRNA metabolic process, Biopolymer metabolic process, G1/S transition of mitotic cell cycle, Cellular carbohydrate metabolic process		mean: 7.4%, Std. Dev.: 2.6

Table 3.2: Cellular state specific protein interaction networks. Metabolic phase, early 1,2 ; early mid, middle, mid-late and late phases represent different phases of sporulation. Early and late G1, G2, M and S phases are different phases of cell cycle. In the second column, we have shown some of the proteins present in the corresponding network. Some of the most enriched GO terms found in the reconstructed networks are shown in the third column. The fourth column shows the mean histograms of cyclic structures for each reconstructed network. The 'X' axis represents the the cycle lengths, the 'Y' axis represents the number of cycles in log scale(zero values are plotted as zeros instead of  $-\infty$  in Log scale). We constructed 2000 protein interaction networks for each cellular state by randomly adding, deleting and rearranging 20% of the interactions of the inferred networks and calculated mean histograms for each state in order to make our inference robust against noise [Yu and Gerstein \(2006\)](#). The last column shows the mean and variance of percentage of proteins which take part in cyclic structures. On an average less than 15% of all the proteins in a particular cellular state form feedback mechanisms. The rest of the proteins operate in a hierarchical manner. The standard deviation of the above quantity is moderately small indicating that our estimation of cyclicity is fairly insensitive to observational and inferential noise.



## Chapter 4

### **Differences in network topologies preferentially occur in acyclic network parts**

In the last two chapters we find that the feedback loops of a biochemical network are more likely to remain conserved than the rest of the network due to their dynamic properties. In this chapter we shall look into the available data to provide evidence in support of our theory. The best way of doing so is to compare the cyclic and acyclic parts of biochemical networks of different yeast species and find out which patterns are more conserved than the others. This is not possible in the current context since the interaction data is not available for most of the yeast species apart from *S. cerevisiae*. Though the genomes of many yeast species are sequenced their interactomes are yet to be determined. However,

#### 4.1 Comparison of genomes interactomes of *S. cerevisiae* and *S. pombe*

---

there are several signalling, cell cycle and metabolic modules which are studied in multiple yeast species such as *S. cerevisiae*, *S. pombe*, *C. albicans*, *C. lusitaniae*, *C. glabrata*, *K. lactis*. We shall consider the cyclic and the acyclic parts of these modules for our analysis to find out which parts are more conserved than the others. Before going into such analysis we shall look into different aspects of the genomes and interactomes of two of the most well studied yeast species, *S. cerevisiae* and *S. pombe*. This is because the genome of *S. pombe* is fully sequenced and partial data is also available for its interactome.

#### 4.1 Comparison of genomes interactomes of *S. cerevisiae* and *S. pombe*

*Saccharomyces cerevisiae*, has diverged from fission yeast *Schizosaccharomyces pombe*, almost a billion years ago (Peng et al. (2005)). Despite such a large evolutionary distance both yeast species have remarkably similar physiological properties. For example, *S. cerevisiae* usually grows on sugars such as glucose, galactose, fructose, cellulose, sucrose etc and is capable of both aerobic and anaerobic respirations. Ammonia and urea are its main nitrogen sources. It can live in both haploid and diploid form. Fission yeast too has limited aerobic respiration and anaerobic fermentation capabilities on many different sugars. Like budding yeast, fission yeast has both haploid and diploid cell types.

#### 4.1 Comparison of genomes interactomes of *S. cerevisiae* and *S. pombe*

---

On the other hand, these two yeast species are different in many aspects, such as:

- budding yeast is round to ovoid shaped, fission yeast is rod shaped
- budding yeast is often diploid, fission yeast is often haploid
- budding yeast stays in the G1 phase of cell cycle for an extended period where as fission yeast remains in the G2 phase for an extended period, etc.

The genomes of these two species also carry evidence of their large evolutionary distance. Both species carries genes which are found conserved in higher Eukaryots but are not conserved between themselves. Budding yeast genome is considerably larger than fission yeast. Budding yeast has 5600 open reading frames where fission yeast has approximately 4970. We discuss a comparison of the genomes of budding and fission yeast as found in literature in order to seek evidence to support our theory of biomolecular network evolution.

##### 4.1.1 Genome conservation in budding and fission yeast

The genome of fission yeast is fully sequenced in 2002 at Sanger institute. A comparison of budding and fission yeast genome has been done by Peng et al. (2005). It is observed by Peng et al. (2005) that

#### 4.1 Comparison of genomes interactomes of *S. cerevisiae* and *S. pombe*

---

among 4061 fission yeast genes 2675 (66%) are conserved in budding yeast. The rest of the genes in fission yeast genomes have no ortholog in budding yeast. Among 5052 budding yeast genes 2636(52 %) are found conserved in fission yeast. It's fair to say around 50-70% of yeast genes are conserved even among the most distant yeast species. The genome conservation of budding and fission yeast are shown in figure 4.1(a). However, even the conserved genes may have different functionality due to evolutionary rewiring. For example, Peng et al. (2005)'s study reveals that although almost half of all 800 budding yeast cell cycle regulated genes are conserved in fission yeast, only 142 of the conserved genes are also regulated by cell cycle in fission yeast. Thus, around 256 budding yeast cell cycle regulated genes are found conserved in fission yeast but their fission yeast homologues do not oscillate in correlation with cell cycle. Similarly, among 747 fission yeast cell cycle regulated genes 417 are found conserved in budding yeast but do not oscillate in synchrony with cell cycle. This indicates large differences between the topologies of biochemical networks of these two species. In another study of transcription regulation during copper and iron uptake (Rustici et al. (2007)) by both species it was revealed that conserved gene clusters perform different tasks in different species. These studies suggest that differences in the architecture of topologies of the gene regulatory network exceed genetic conservation. The above argument is further supported by the experiments performed by Dixon et al. (2008) . Rather than comparing genome homology, Dixon et al. (2008) compared the transcription interactions of budding and fission yeasts and found that only 23% of budding yeast interactions are conserved in fission yeast and the rest of the genes are rewired due to evolutionary renovation. The conservation of gene

#### 4.1 Comparison of genomes interactomes of *S. cerevisiae* and *S. pombe*

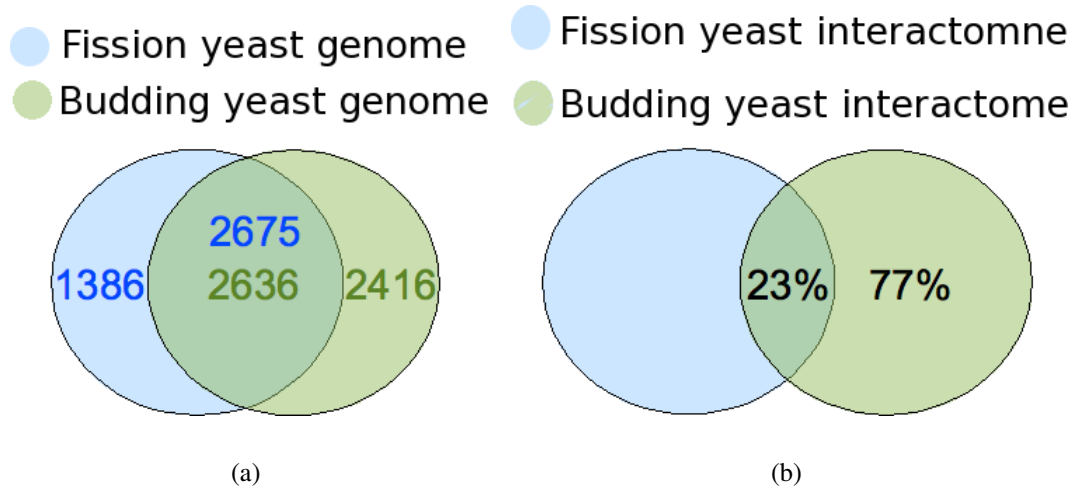


Figure 4.1: Conservation of genomes and gene regulatory networks in budding and fission yeast. (a) shows the number of conserved open reading frames in both species and (b) shows the percentage of genetic interaction conserved in both species.

regulation in budding and fission yeast are shown in figure 4.1(b).

From the above studies it can be concluded that a large portion, approximately 75% of gene regulatory networks of budding and fission yeast have different topologies which arise from extensive evolutionary renovation for almost a billion years. Since their gene regulatory networks have gone through extensive evolutionary rewiring it is logical to predict similar phenomena for their protein interaction networks. However, a statistical comparison of protein interaction networks of six different species, namely, *E. coli*, *H. pylori*, *S. cerevisiae*, *C. elegans*, *D. melangoster*, *M. musculus* by Liang et al. (2006) confirms that despite relatively high level of sequence conservation in proteins of different species their interaction patterns have altered extensively. According to Liang et al. (2006), such alteration may or may not re-allocate differ-

ent protein complexes to different functional groups. Though in many cases changes in interaction patterns altered the function of homologous proteins in different species, in some cases it is also found that despite reshuffling their interaction patterns a group of proteins retain their functionality in many different species (Liang et al. (2006)). In other words, evolutionary renovations altered the interactions patterns of even the conserved proteins both globally and within the same modules. We presume that the global rewiring of protein protein interactions must not result in instability in the dynamic behaviour of a protein interaction network, additionally, the local intra-modular rewiring of PPIs must not violate the stability of the corresponding module in order for such rewiring to sustain in future progeny.

The studies of Dixon et al. (2008); Liang et al. (2006) suggest that most of the molecular interaction networks of budding and fission yeast have dissimilar network topologies. Since, most parts of yeast biochemical networks are acyclic it's evident that most evolutionary renovations take place in the acyclic parts of the network. But this is only indirect and anecdotal evidence to support our hypothesis that the evolutionary renovations preferentially occur in the acyclic parts of biochemical networks. The above studies do not provide conclusive evidence to suggest that the feedback loops in the biochemical networks of yeast are better conserved than their acyclic counterparts. Hence, in the next section we shall discuss the evolutionary conservation of metabolic pathways, cell cycle and signalling modules in different yeast species. Among all different types of pathways metabolic networks are the most widely studied in many different yeast species and high quality data are available for

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

---

these networks. So, we shall first investigate the conservation of network structures of the metabolic networks of different yeast species.

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

In order to investigate the evolution of network topologies in yeast metabolic networks we downloaded the integrated metabolic networks of thirteen different yeast species namely, *C. albicans*, *C. dubliniensis*, *C. glabrata*, *C. tropicalis*, *D. hansenii*, *K. lactis*, *L. elongisporous*, *L. thermotolerans*, *P. pastoris*, *S. cerevisiae*, *S. pombe*, *V. polyspora* and *Z. rouxii* from KEGG pathway database <sup>1</sup>. Each pathway is then transformed into a set of biochemical reactions following the KEGG reaction database <sup>2</sup>. Each set of reactions is then converted into a directed graph by drawing edges from reactants to products since metabolic flux flows in the direction of reactant to product. The topologies of the resulting metabolic networks are then compared with each other. In order to find out whether the topologies of the acyclic or the cyclic parts are evolutionarily more conserved we separated the cyclic parts of the networks from the acyclic parts. We then determined the similarities between the cyclic parts of the metabolic networks of different yeast species with each other and the acyclic parts of the metabolic networks with each other. The similarities between the cyclic structures are measured by

---

<sup>1</sup><http://www.genome.jp/kegg/pathway.html>

<sup>2</sup><http://www.genome.jp/kegg/reaction/>

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

---

finding % of edges common between the cyclic structures of two different yeast species. For example, the conservation of the cyclic structures of the metabolic network of *C. albicans* in *S. cerevisiae* is calculated as  $C_{CAL-SCE}^{cyc} = \frac{100E_{CAL-SCE}^{cyc}}{E_{CAL}^{cyc}}$ , where,  $E_{CAL-SCE}^{cyc}$  is the number of common edges found in the cyclic parts of the metabolic networks of *C. albicans* and *S. cerevisiae* and  $E_{CAL}^{cyc}$  are the total number of edges in the cyclic parts of the metabolic network of *C. albicans*. The conservation of the acyclic parts of the metabolic networks of different yeast species are also calculated in a similar manner. For example,  $C_{CAL-SCE}^{acyc} = \frac{100E_{CAL-SCE}^{acyc}}{E_{CAL}^{acyc}}$ . Finally, for each pair of yeast species the conservation score of the cyclic parts are compared with the conservation score of the acyclic parts. For example, when investigating the evolution of the topology of metabolic networks of *C. albicans* and *S. cerevisiae* we compare  $C_{CAL-SCE}^{cyc}$  and  $C_{SCE-CAL}^{cyc}$  with  $C_{CAL-SCE}^{acyc}$  and  $C_{SCE-CAL}^{acyc}$ . If  $C_{CAL-SCE}^{cyc} > C_{CAL-SCE}^{acyc}$  and  $C_{SCE-CAL}^{cyc} > C_{SCE-CAL}^{acyc}$  we conclude that the cyclic structures of the metabolic networks are more conserved between *C. albicans* and *S. cerevisiae* compared to the acyclic structures. Since we have data for 13 different yeast species we have 169 comparison scores for both cyclic and acyclic structures, 13 of which are self comparisons, i.e.  $C_{CAL-CAL}^{cyc}$  and  $C_{CAL-CAL}^{acyc}$  etc. The results of the analysis of conservation scores of the cyclic and acyclic parts of yeast metabolic network is shown in figure 4.2.

From figure 4.2 it can be noticed that there is a clear difference between the conservation of the cyclic and acyclic structures of the metabolic networks of different yeast species. Evidently, in most cases each block of figure 4.2(a) is lighter than the corresponding block of fig-



## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

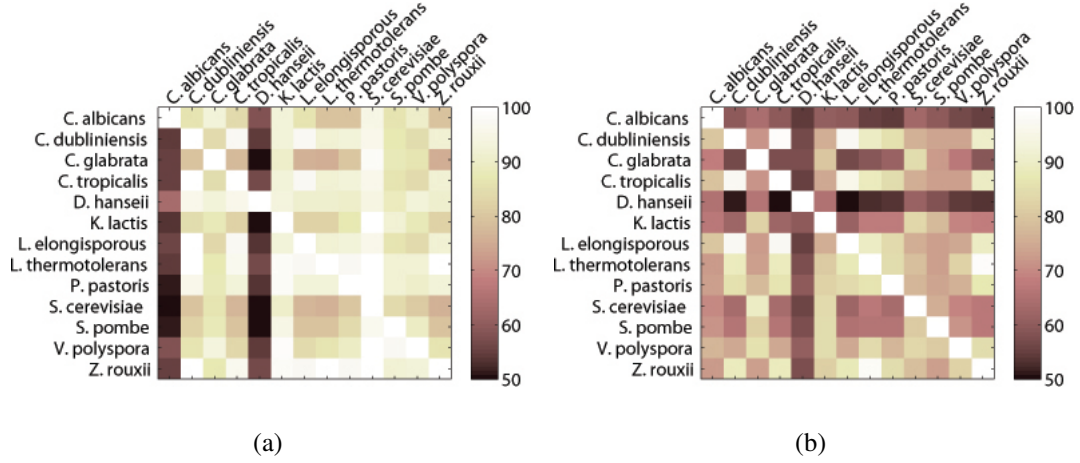


Figure 4.2: Conservation of network structures in cyclic (a) and acyclic (b) parts of yeast metabolic network. Each block represents the conservation score of the cyclic (a) and acyclic structure (b) of the corresponding species. For example, the third block in the second row of figure (a) maps the conservation score  $C_{CDU-CGL}^{cyc}$  in the colour space shown on the right of each figure. Figure (a) is visibly lighter than figure (b), except from first and fifth column, suggesting, generally, higher conservation in the network structures of the cyclic parts of yeast metabolic pathways. The first and fifth column of both figure (a) and (b) suggests the opposite for *C. albicans* and *D. hansenii*. The first(fifth) column of figure (a) suggests that only around (50-60%) of the cyclic structures found in the other yeast species are also found present in *C. albicans*(*D. hansenii*). But a closer look at the first(fifth) row suggest that (80-90%) of the cyclic structures found in the *C. albicans*(*D. hansenii*) are also present in other yeast species. This can only be possible if there are fewer cyclic structures in the metabolic networks of *C. albicans* and *D. hansenii* compared to the other species. Most of the cyclic structures that are present in the metabolic networks of *C. albicans* and *D. hansenii* are also present in the other yeast species but since the other species have many more cycles in their metabolic networks the opposite is not true. The anomaly in the first and fifth column of figure (b) can be explained in a similar way.

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

---

ure 4.2(b) suggesting higher conservation of cyclic structures compared to the acyclic parts. However, there are two exceptions. The first and fifth columns of figure 4.2(a) are darker than the first and fifth columns of figure 4.2(b) suggesting, apparently the opposite for *C. albicans* and *D. hansenii*, i.e.  $C_{...-CAL}^{acyc} > C_{...-CAL}^{cyc}$  and  $C_{...-DHA}^{acyc} > C_{...-DHA}^{cyc}$ . A closer look at the first and the fifth row of both figure 4.2(a) and 4.2(b) reveals that  $C_{CAL-...}^{cyc} > C_{CAL-...}^{acyc}$  and  $C_{DHA-...}^{cyc} > C_{DHA-...}^{acyc}$ . This anomaly results from the fact that *C. albicans* and *D. hansenii* has smaller number of cyclic structures (according to the KEGG pathway database) compared to the other yeast species. Hence only a fraction of the cyclic structures that are found in the other yeast species are also found conserved in the metabolic networks of *C. albicans* and *D. hansenii*. On the other hand, the rows of figure 4.2(a) and 4.2(b) corresponding to *C. albicans* and *D. hansenii* suggest that most of the feedback loops that are found in the metabolic networks of *C. albicans* and *D. hansenii* are also found conserved in other yeast species. Hence the above anomaly is not in violation of our hypothesis.

Most of the biochemical reactions regarding metabolic networks are enzymatic reactions. Hence the above results naturally give rise to a different question: whether the enzymes which take part in the feedback reactions are better conserved compared to those which take part in acyclic chains. To investigate this possibility we carry out two different experiments. Firstly, we calculate what percentage of enzymes which take part in the cyclic and acyclic parts of the metabolic network have ortholog in other yeast species. Here, we call two enzymes orthologous if the Smith-Waterman-Gotoh similarity scores between them is

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

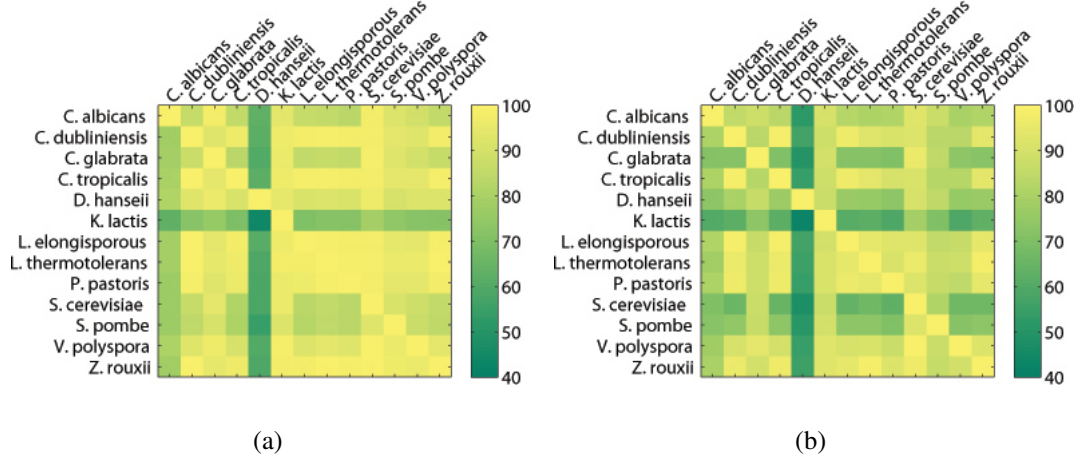


Figure 4.3: : Conservation of enzyme sequences in the cyclic(a) and acyclic(b) parts of the metabolic networks of different yeast species. Each small square in figure (a) and (b) maps the  $Os_{species1-species2}^{cyc}$  and  $Os_{species1-species2}^{acyc}$  to the colour space shown at the right. It can be seen that each small square in figure (a) is visibly lighter than the corresponding square of figure(b) suggesting that higher percentage of enzymes which take part in feedback interactions have orthologs in other species compared to those which take part in acyclic reactions chains.

greater than 100 as suggested in KEGG orthology database. This percentage is calculated as  $Os_{species1-species2}^{cyc/acyc} = \frac{100Enz_{species1-species2}^{cyc/acyc}}{Enz_{species1}^{cyc/acyc}}$ , where  $Enz_{species1-species2}^{cyc/acyc}$  represents the number of enzymes that are found in the cyclic/acyclic parts of the metabolic network of species1 and have at least one ortholog in species2.  $Enz_{species1}^{cyc/acyc}$  represents the total number of enzymes found in the cyclic/acyclic parts of the metabolic network of species1. The  $Os_{species1-species2}^{cyc/acyc}$  scores are shown in figure 4.3. It can be seen from figure 4.3 that a higher percentage of the enzymes which take part in cyclic interaction patterns have orthologs in other species indicating a higher sequence conservation compared to the enzymes which take part in acyclic reaction chains.

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

---

We also investigated whether there is a statistically significant difference between the sequence conservation of the enzymes which take part in feedback loops and those which take part in acyclic reaction chains. Let us denote the set of enzymes of the  $i^{th}$  species by  $\eta_i = \{s_{ij} : i \in I, j = 1 \dots N_i\}$ , here  $N_i$  is the number of enzymes present in the  $i^{th}$  species, and  $s_{ij}$  is the amino acid sequence of the  $j^{th}$  enzyme of the  $i^{th}$  species. Each  $\eta_i$  has two subsets  $\eta_i^c$ , the set of enzymes which take part in the cyclic structures of the metabolic network of the  $i^{th}$  species and  $\eta_i^a$ , the set enzymes which take part in its acyclic interactions. Hence,  $\eta_i = \{\eta_i^c \cup \eta_i^a\}$ . For each enzyme  $s_{ik} \in \eta_i$  we looked for its ortholog in other yeast species by its relevant 'ko' number (Kanehisa et al., 2008). Let us assume that the  $k^{th}$  enzyme  $s_{ik}$  of the  $i^{th}$  species have a set of orthologs  $\eta_j^{ik} = \{s_{jl}^{ik} : l = 1 \dots n_{ijk}\} \subset \eta_j$  in the  $j^{th}, i \neq j$  yeast species. We computed a set of distance measures between the amino acid sequences of  $s_{ik}$  and its orthologs in the  $j^{th}$  species. We denote this set by  $\delta_j^{ik} = \{d_{jl}^{ik} : l = 1 \dots n_{ijk}\}$ . Distances between two amino acid sequences are calculated using the following formula.

$$d_{jl}^{ik} = (1 - \frac{S_{kl}}{S_{kk}})(1 - \frac{S_{kl}}{S_{ll}}) \quad (4.1)$$

where  $S_{kl}$  is the sequence similarity score between the enzyme sequences  $s_{ik}$  and  $s_{jl}^{ik}$ , which is calculated using Smith Waterman Gotoh sequence alignment algorithm. Clearly,  $0 \leq d_{jl}^{ik} \leq 1$ . If the two sequences have close resemblance to each other then  $d_{jl}^{ik} \approx 0$  and in the opposite case  $d_{jl}^{ik} \approx 1$ . If the distance between two sequences is high then it assume that the corresponding enzyme have evolved at a higher rate and vice versa. However, if  $\eta_j^{ik} = \emptyset$  we assign  $\delta_j^{ik} = \{1\}$ . Doing so takes into account the

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

---

fact that enzyme  $s_{ik}$  has no ortholog in the  $j^{th}$  yeast species. This reduces statistical bias in our study. We constructed two large sets of distance measures between amino acid sequences  $\Delta^c = \{\cup_{i,j \in I; i \neq j} \cup_{\forall s_{ik} \in \eta_i^c} \delta_j^{ik}\}$  and  $\Delta^a = \{\cup_{i,j \in I; i \neq j} \cup_{\forall s_{ik} \in \eta_i^a} \delta_j^{ik}\}$ . Hence,  $\Delta^c$  is the set of distance measures of the enzymes which take part in feedback interactions from their orthologs and  $\Delta^a$  is the same for the enzymes which take part in acyclic interactions. We carried out statistical comparisons of  $\Delta_c$  and  $\Delta_a$  to investigate whether the rates of evolution in feedback enzymes are significantly different to those which take part in acyclic interactions. We estimated the probability density functions (pdf) of  $\Delta_c$  and  $\Delta_a$  using kernel density estimation<sup>1</sup>. The pdfs are shown in figure 4.4. It can be seen from figure 4.4 that the pdf of the rate of evolution of the enzymes which take part in the cyclic interactions,  $pdf(d_c)$  has a sharper peak near 0 compared to  $pdf(d_a)$ . This suggests that a larger fraction of the enzymes which take part in the cyclic structures evolves at a slower rate.

We further investigated whether the distributions of  $\Delta_a$  and  $\Delta_c$  are statistically significantly different. As the most primitive statistical measures we calculated the means and medians of  $\Delta^c$  and  $\Delta^a$ . The means and medians of the above samples are  $mean(d_c) \approx 0.3$  and  $median(d_c) \approx 0.17$ ,  $mean(d_a) \approx 0.4$  and  $median(d_a) \approx 0.27$ . This suggests that the average distances between feedback loop related enzymes and their orthologs are smaller compared to the acyclic interaction related enzymes, i.e., the feedback related enzymes have a slower rate of evolution. However, a more systematic way of comparing  $\Delta^c$  and  $\Delta^a$  is to do statistical hypothesis tests and confirm that they come from

---

<sup>1</sup>ksdensity() function in MATLAB

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

---

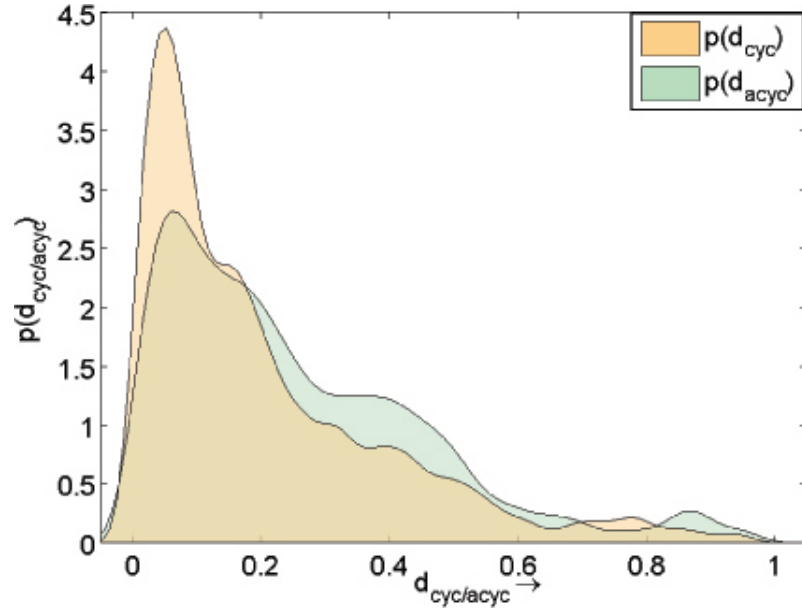


Figure 4.4: PDF of rate of evolution in the conserved enzymes (Smith-Waterman-Gotoh sequence similarity score  $>100$ ) found in cyclic and acyclic parts of the metabolic networks of all 13 yeast species. The rate of evolution is represented by evolutionary distance calculated from the similarity scores and ranges from 0-1. Enzymes whose orthologs have an evolutionary distance closer to 0 evolve at a slower rate and vice versa. The above picture shows that the probability density function of the rate of evolution of the enzymes which take part in cyclic interactions has a considerably sharper peak near zero compared to that of the acyclic parts. This suggests that a larger fraction of the enzymes which take part in the cyclic reactions of the metabolic networks of different yeast species evolve in a slower rate.

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

---

different distributions. Kolmogorov Smirnov (KS) test is one of the most commonly used hypothesis test in similar scenario. In our case, a KS test suggest that  $p(d_c) \neq p(d_a)$  with very high certainty  $p = 0$ . But, a classical hypothesis test such as KS test may become overpowered due to large number of samples in  $\Delta^c$  and  $\Delta^a$  and may not be reliable. Hence, we resort to a Bayesian alternative to KS test (Wolpert, 1995). It calculates a distance measure between the two data samples. Let us consider that  $\Delta^c$  comes from a distribution  $p(d_c)$  and  $\Delta^a$  comes from  $p(d_a)$  and the parameters of the distributions  $p(d_c)$  and  $p(d_a)$  are  $\pi_c$  and  $\pi_a$ . The distance between the samples  $\Delta^c$  and  $\Delta^a$  is calculated using the following formula which is taken from (Wolpert, 1995).

$$\begin{aligned}
 E(\Sigma_{ac}) &= \int (\pi_c - \pi_a)^2 p(\pi_c, \pi_a | \Delta^c \Delta^a) d\pi_c d\pi_a \\
 &= \frac{J}{K} \\
 J &= \sum_{i=1}^m K(h_c + 2(i), h_a) + K(h_c, h_a + 2(i)) - 2K(h_c + 1(i), h_a + 1(i)) \\
 K &= K(h_c, h_a) = \frac{\prod_{i=1}^m \Gamma(h_c(i) + 1) \Gamma(h_a(i) + 1)}{\Gamma(N_1 + m) \Gamma(N_2 + 1)}
 \end{aligned} \tag{4.2}$$

where  $p(\pi_c, \pi_a | \Delta^c \Delta^a)$  is calculated using the Bayes formula, i.e.,

$$p(\pi_c, \pi_a | \Delta^c \Delta^a) = \frac{p(\Delta^c, \Delta^a | \pi_c, \pi_a) p(\pi_c, \pi_a)}{\int p(\Delta^c, \Delta^a | \pi_c, \pi_a) p(\pi_c, \pi_a) d\pi_c d\pi_a} \tag{4.3}$$

$h_c$  and  $h_a$  are the histograms of  $\Delta^c$  and  $\Delta^a$  with  $m$  numbers of bins.  $h_c + 2(i)$  denotes the histogram  $h_c$  with an extra 2 added to its  $i^{th}$  bin.

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

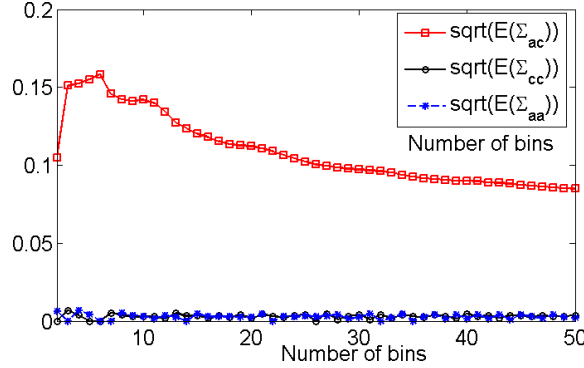


Figure 4.5: A Bayesian distance measure between the probability distributions of  $\Delta_a$  and  $\Delta_c$ . The expected distances between the distributions of  $\Delta_a$  and  $\Delta_c$ , denoted by  $\sqrt{E(\Sigma_{ac})}$ , are plotted against the bin sizes  $m = 2 \dots 50$ .  $\sqrt{E(\Sigma_{aa})}$  and  $\sqrt{E(\Sigma_{cc})}$  are also plotted which represents the expected distance between two samples drawn from the same distribution. It can be seen from this figure that  $\sqrt{E(\Sigma_{ac})} > \sqrt{E(\Sigma_{aa/cc})} \forall m = 2 \dots 50$ .

$\Gamma()$  is the gamma function and  $h_c(i)$  is the value of the  $i^{th}$  bin of  $h_c$ . The formula of equation 4.3 is derived assuming that the samples  $\Delta^a$  and  $\Delta^c$  come from multinomial distributions with Dirichlet priors. The distance measure  $E(\Sigma_{ac})$  depends on the number of bins  $m$ . Hence we calculated  $E(\Sigma_{ac})$  for  $m = 2 \dots 50$ . Additionally, we generated two data samples  $R_{\Delta^c}$  and  $R_{\Delta^a}$  by resampling the distributions  $p(d_c)$  and  $p(d_a)$ .  $R_{\Delta^c}$  and  $R_{\Delta^a}$  have the same number of data samples as  $\Delta^c$  and  $\Delta^a$ . We computed the distances between  $R_{\Delta^c}$  and  $\Delta^c$ ,  $R_{\Delta^a}$  and  $\Delta^a$ , using the formula shown in equation (1) and denoted them by  $E(\Sigma_{cc})$  and  $E(\Sigma_{aa})$ . We used  $E(\Sigma_{cc})$  and  $E(\Sigma_{aa})$  as reference distance between two data samples if they were drawn from the same distributions. We have plotted  $\sqrt{E(\Sigma_{ac})}$ ,  $\sqrt{E(\Sigma_{cc})}$  and  $\sqrt{E(\Sigma_{aa})}$  for  $m = 2, 3, \dots 50$  in Fig. 4.2. It can be seen from Fig. 2I that  $E(\Sigma_{ac}) \gg E(\Sigma_{cc/aa}), \forall m \in 2, 3 \dots 50$  suggesting significant difference between the distributions of  $\Delta^c$  and  $\Delta^a$ .



## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

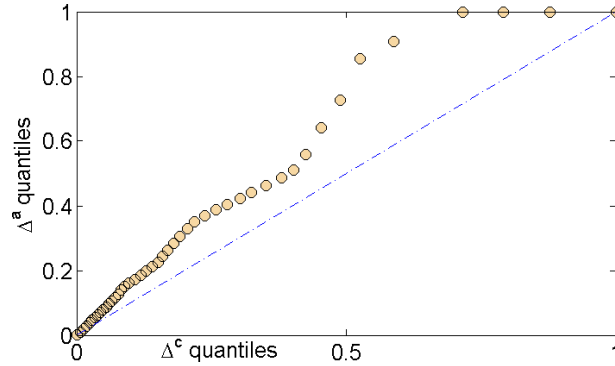


Figure 4.6: Quantile-quantile plot of  $\Delta_a$  and  $\Delta_c$ .

The above statistical test tells us that  $\Delta^c$  and  $\Delta^a$  come from different distributions. To understand the qualitative statistical difference between these samples we plotted a quantile-quantile (q-q) plot of the two samples in Fig. 4.2. It can be seen from fig. 4.2 that the q-q plot has a significant upward deviation from the  $45^\circ$  reference line suggesting that  $q_i(\Delta^a) > q_i(\Delta^c) \forall i$ , where  $q_i(\Delta^{a/c})$  is the  $i^{th}$  quantile of  $\Delta^{a/c}$ . This suggests that a larger fraction of  $\Delta^c$  has smaller values than  $\Delta^a$  which essentially suggests that the enzymes of the cyclic interactions evolve slowly compared to the enzymes of acyclic interactions.

As an example of the conservation of cyclic network structures and rewiring of acyclic reaction paths in yeast metabolic pathways we present a comparison of the phospholipid biosynthesis pathway of *S. cerevisiae* and *C. albicans* in Fig. 4.7<sup>1</sup>. We found that the feedback mechanisms of this pathway are highly conserved in both species, whereas the acyclic interaction sequences have changed.

<sup>1</sup><http://www.genome.jp/kegg/pathway.html>

## 4.2 Differences in network topologies preferentially occur in the acyclic parts of yeast metabolic networks

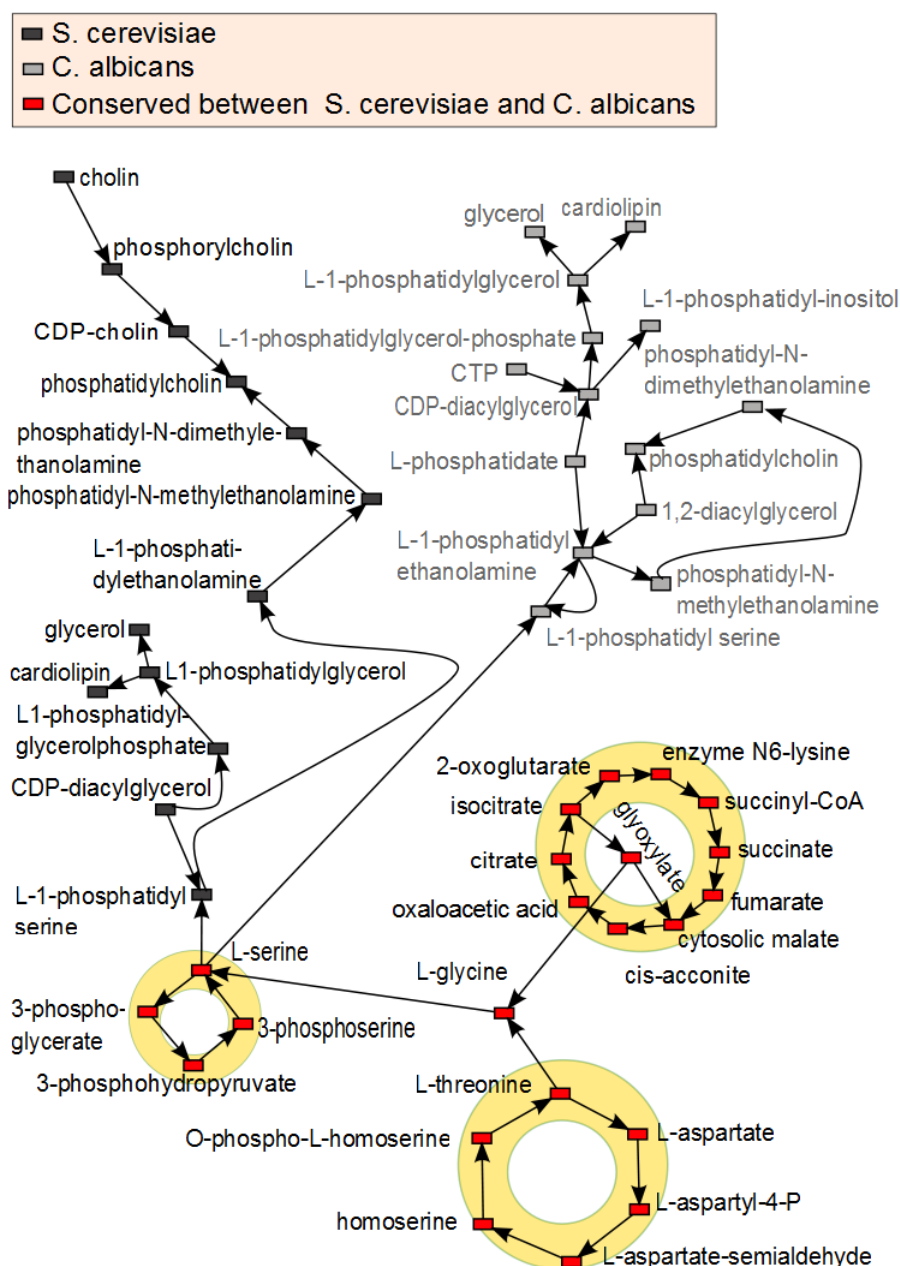


Figure 4.7: Differential evolution in the cyclic and acyclic parts of yeast metabolic network. In this figure we compare phospholipid biosynthesis pathways of *S. cerevisiae* and *C. albicans*. The feedback mechanisms responsible for the glyoxylate cycle, serine and threonine biosynthesis are conserved in both species. Main differences arise in its regulation by exogenous precursors, inositol and choline, which is governed by the acyclic parts of the pathway that is responsible for the biosynthesis of lipids. The red nodes represent the compounds which are parts of conserved cyclic modules present in the metabolic networks of both species. The acyclic parts of this pathway are represented by dark (*S. cerevisiae*) and light gray (*C. albicans*) nodes. The Kennedy pathways are not explicitly shown.

### **4.3 t *S. cerevisiae* genes which take part in biochemical feedback interactions evolve slowly**

---

The above results provide reasonable evidence that the feedback mechanisms of the metabolic networks of yeast are better conserved than the acyclic reaction chains. It is also found that the enzymes which form these feedback loops also evolve at a slower rate. This raises the question whether the above results hold true for non-metabolic pathways, such as gene regulatory networks, signalling and cell cycle related pathways etc. We attempt to answer this question in the next section.

### **4.3 t *S. cerevisiae* genes which take part in biochemical feedback interactions evolve slowly**

To understand how the topologies of biochemical networks evolve one need to compare the biochemical networks of several species. Among all the yeast species only *S. cerevisiae* has a near complete interactome. The unavailability of network data for other yeast species renders such studies infeasible. Hence we are left with little choice but to investigate the evolution of the genes which take part in feedback interactions in the biochemical network of *S. cerevisiae*. If we find that these genes are better conserved compared to average *S. cerevisiae* genome then at least we can provide evidence to support some of our claims.

We manually compiled a list of 753 genes in *S. cerevisiae* which are found to take part in feedback interactions in *S. cerevisiae* gene regulatory, signalling, cell cycle and metabolic networks from avail-

### 4.3 t *S. cerevisiae* genes which take part in biochemical feedback interactions evolve slowly

---

able literature. A detailed description of the data, and relevant literature references are provided in supplementary table ???. Let us denote the set of amino acid sequences related to the *S. cerevisiae* genome by  $\eta_i = \{s_{ik} : k = 1 \dots N_i, i = SCE\}$  and the sequences of the genes which form feedback loops in the biochemical network of *S. cerevisiae* by  $\eta_i^c$ . For each gene  $s_{ik}$ , where  $i = SCE$ , we looked for its orthologs (the genes with same 'ko' number in KEGG orthology database [Kanehisa et al. \(2008\)](#)) in other yeast species. Let us denote the set of orthologs of gene  $s_{ik}$  in the  $j^{th}$  yeast species by  $\eta_j^{ik}, i = SCE, j \in I, j \neq i$ . For each  $s_{ik}$  we calculated the set of evolutionary distance  $\delta_j^{ik}$  between  $s_{ik}$  and its orthologs  $\eta_j^{ik}$  as described in the previous section. If  $\eta_j^{ik} = \emptyset$  we assign  $\delta_j^{ik} = \{1\}$ . Finally we constructed the set  $\Delta^t = \cup_{j \in I, j \neq SCE} \cup_{s_{ik} \in \eta_i, i = SCE} \delta_j^{ik}$  which represents the set of distances of the yeast genes from their orthologs found in other species. Similarly, we constructed a similar set of distances for the genes which form feedback loops in *S. cerevisiae* and their orthologs in other yeast species. We denote this set by  $\Delta^l$ . We then carried out statistical comparison of  $\Delta^t$  and  $\Delta^l$  to investigate the qualitative and quantitative differences between these samples. Firstly, as the simplest statistical measure we calculated the mean and median of  $\Delta^l$  and  $\Delta^t$ ,  $mean(\Delta^l) = 0.52$ ,  $median(\Delta^l) = 0.44$ ,  $mean(\Delta^t) = 0.77$ ,  $median(\Delta^t) = 1$ , Since  $mean(\Delta^l) < mean(\Delta^t)$ , it can be assumed that on an average the genes which form feedback loops are better conserved compared to an average gene in *S. cerevisiae* genome. For more systematic statistical analysis, we calculated the difference between the probability distributions of  $\Delta^t$  and  $\Delta^l$  using the formula shown in equation (4.2) for multiple bin sizes  $m = 2 \dots 50$ . As stated in the previous

### 4.3 *S. cerevisiae* genes which take part in biochemical feedback interactions evolve slowly

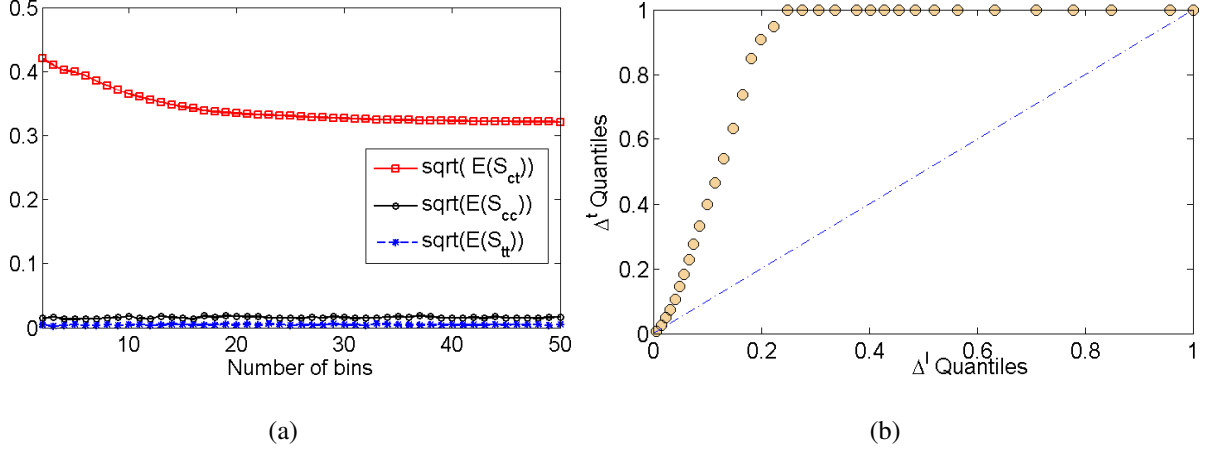


Figure 4.8: Differential evolution in the amino-acid sequences of the genes related to the feedback mechanisms of *S. cerevisiae* biochemical networks. (A) shows the Bayesian distance measures between  $\Delta^l$  and  $\Delta^t$ . (B) shows the quantile-quantile plot of  $\Delta^l$  and  $\Delta^t$

section, we generated two samples,  $R_{\Delta^l}$  and  $R_{\Delta^t}$  by re-sampling the distributions of  $\Delta_l$  and  $\Delta_t$  and calculated the differences between  $\Delta^t$  and  $R_{\Delta^t}$ , denoted by  $E(\Sigma_{tt})$ , and,  $\Delta^l$  and  $R_{\Delta^l}$  denoted by  $E(\Sigma_{ll})$ . We used  $E(\Sigma_{tt})$  and  $\Sigma_{ll}$  as reference. We have shown  $\sqrt{E(\Sigma_{lt})}$ ,  $\sqrt{E(\Sigma_{tt})}$  and  $\sqrt{E(\Sigma_{ll})}$  in Fig. 4.8(a). It can be seen from Fig. 4.8(a) that  $\sqrt{E(\Sigma_{lt})} \gg \sqrt{E(\Sigma_{ll/tt})}, \forall m \in 2 \dots 50$ , which suggests that distributions of  $\Delta^l$  and  $\Delta^t$  have significant statistical difference. A quantile-quantile plot of  $\Delta^l$  and  $\Delta^t$  is shown in Fig. 4.8(b). The stiff upward deviation of the qq-plot from the 45 degree reference line suggests that the average distance of the *S. cerevisiae* genome from other yeast species is much higher compared to the genes which form feedback mechanisms, i.e. the feedback related genes are better conserved in other species.

Though there is a lack of data on the biochemical networks of yeast

species other than *S. cerevisiae*, many specific pathways regarding important physiological mechanisms of yeast have been studied in different yeast species. In the next section, we use available data from literature to provide anecdotal evidence of differential evolution of the topology of cyclic and acyclic parts of yeast biochemical networks, especially signalling, cell cycle and other transcription regulatory networks.

#### 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

The best way to provide evidence to support our hypothesis is to execute a systematic statistical comparison of the biochemical networks of different yeast species. Though the genomes of many yeast species are sequenced, their interactome is yet to be determined. In other words, there are not enough interaction data for yeast species other than *S. cerevisiae* to carry out a statistical comparison of their conserved and non conserved modules. Therefore, we resort to case studies for the modules for which data are available for at least two different yeast species. Some examples are as follows.

The regulatory negative feedback interactions of HOG1, SKO1 and PTP3 in *S. cerevisiae* is also conserved in its distant relative *S. pombe* in the form of a feedback made of STY1, ATF1, PYP2 (Fig. 4.9) [Proft et al. \(2005\)](#); [Wilkinson et al. \(1996\)](#). However, the orthologs SKO1

#### 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

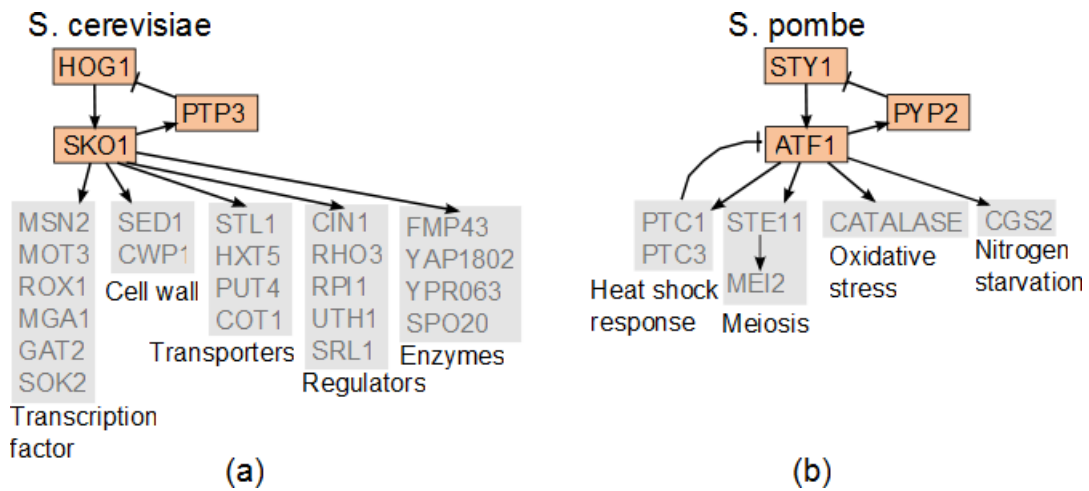


Figure 4.9: Conserved regulatory feedback module in the osmoregulation pathways of *S. cerevisiae* and *S. pombe*. Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey.

and ATF1 regulate different sets of genes in different species (Proft et al. (2005); Wilkinson et al. (1996)). This is likely because *S. cerevisiae* has different signalling and regulatory pathways for different types of stress responses (Wilkinson et al. (1996)) and the above module only deals with osmotic stress. Hence, the genes related to heat shock, oxidative stress, nitrogen stress etc. are not regulated by SKO1. On the other hand, *S. pombe* shares the same signalling and regulatory pathway for many kinds of stress. Hence, ATF1, the *S. pombe* ortholog of SKO1 regulates genes that deal with many different kinds of stress (Davidson et al. (2004); Nguyen and Shiozaki (1999)) including oxidative stress, heat shock, and nitrogen scarcity. This is an illustrative example of regulatory modules where the feedback loops are conserved but the surrounding network of regulators and effector has undergone extensive evolutionary rewiring.

The transcriptional negative feedback mechanism mediated by MET30

#### 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

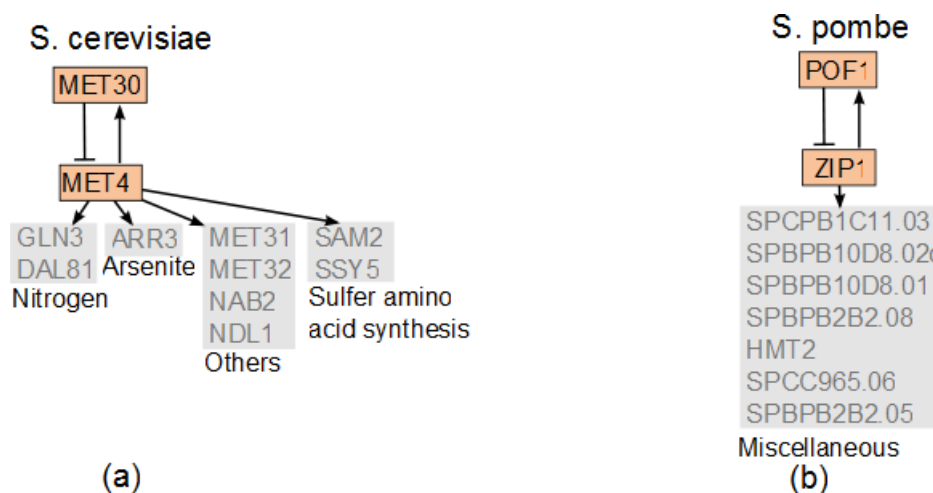


Figure 4.10: Conserved regulatory feedback module in the cadmium response pathways of *S. cerevisiae* and *S. pombe*. Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey.

and MET4 in the cadmium response module of yeast *S. cerevisiae* is conserved in *S. pombe* in the form of a negative feedback mediated by POF1 and ZIP1 (Harrison et al. (2005); Jonkers and Rep (2009)) (Fig. 4.10). *S. cerevisiae* MET4 is found to regulate genes which are involved in nitrogen starvation, sulfur amino acid synthesis, arsenite response etc. whereas the genes regulated by *S. pombe* ortholog ZIP1 of *S. cerevisiae* MET4 are not found to take part in these events (Harrison et al. (2005); Jonkers and Rep (2009)). This module is also another example of a regulatory unit where the feedback modules are conserved but the surrounding regulatory setting have evolved.

The transcriptional feedback loops mediated by SBF, YOX1 and SWI4 that promotes G1-S transition (Fig. 4.11) are conserved in *S. pombe* in the form of feedback loops made of MBF, YOX1 and CDC10



#### 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

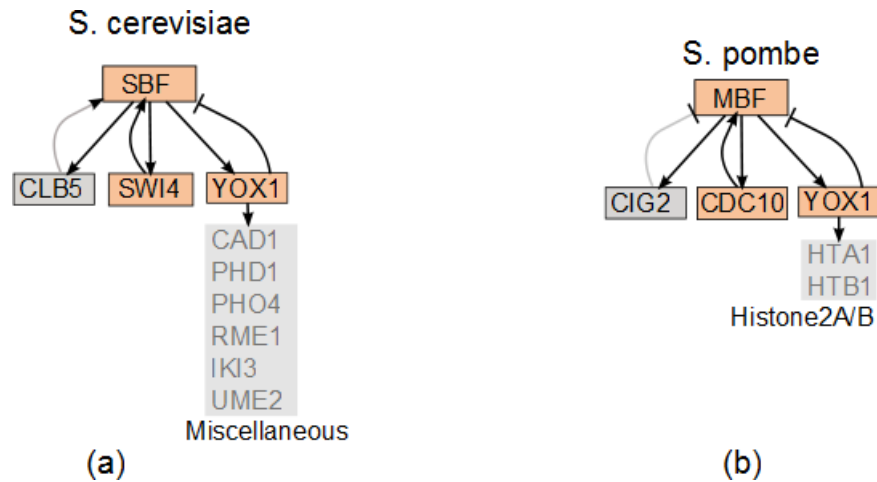


Figure 4.11: Conserved regulatory feedbacks in the cell cycle modules of *S. cerevisiae* and *S. pombe*. Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey. CLB5 and CIG2 are shown in grey because the feedback loop mediated by them are not conserved in both the species. CLB5 mediates a positive feedback loop with SBF in *S. cerevisiae* whereas CIG2, the *S. pombe* ortholog of CLB5 mediates a negative feedback loop with MBF, the *S. pombe* ortholog of SBF. The rest of the feedback loops in this modules are conserved in exact form.

(Aligianni et al. (2009); Horak et al. (2002); Raithatha and Stuart (2005)).

We also look into the set of genes which are regulated by the YOX1 in both *S. cerevisiae* and *S. pombe*. YOX1 regulates PHD1 and PHO4 in *S. cerevisiae* which control pseudohyphal growth and phosphate genes respectively. A similar mechanism is not found in *S. pombe* (Aligianni et al. (2009); Horak et al. (2002); Raithatha and Stuart (2005)). On the other hand, *S. pombe* YOX1 regulates histone 2A and B genes which are not found regulated by *S. cerevisiae* YOX1 (Aligianni et al. (2009); Horak et al. (2002); Raithatha and Stuart (2005)).

GAL genes mediate feedback loops which are involved in galactose metabolism are found to be partly conserved between *S. cerevisiae* and

#### 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

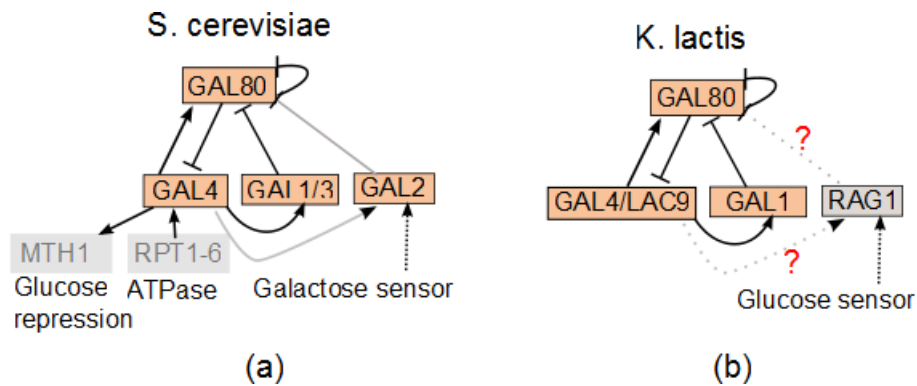


Figure 4.12: Conserved regulatory feedback module in the galactose metabolism pathways of *S. cerevisiae* and *K. lactis*. Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey. RAG1 is shown in grey because the feedback loop between GAL80, GAL4 and RAG1 might not be present in *K. lactis*, because, unlike GAL2, the *S. cerevisiae* homolog of RAG1, it does not act as a galactose sensor, rather it act as a glucose transporter in *K. lactis*.

*K. lactis*. The feedback loops made of GAL4, GAL80 and GAL1/3 in *S. cerevisiae* are found conserved in *K. lactis* in the form of feedback loops between GAL80, GAL4/LAC9 and GAL1 (Fig. 4.12) (Anders et al. (2006); Bhat (2008); Zenke et al. (1993)). One of the principle component of this module, GAL4 regulates different sets of genes in *S. cerevisiae* and *K. lactis*. GAL4 in *S. cerevisiae* is found to regulate glucose repression genes and is also regulated by genes which promote ATPase synthesis. These activities are not detected in *S. pombe* for its GAL4 homolog (Anders et al. (2006); Bhat (2008); Zenke et al. (1993)).

Another feedback module mediated by coupled negative feedbacks made of RPN4, YAP1 and PDR1 modulate drug resistance in *S. cerevisiae*. This module is found to be conserved in *C. glabrata* in the form of a feedback module mediated by RPN4, CGAP1, PDR1 (Fig. 4.13) (

#### 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

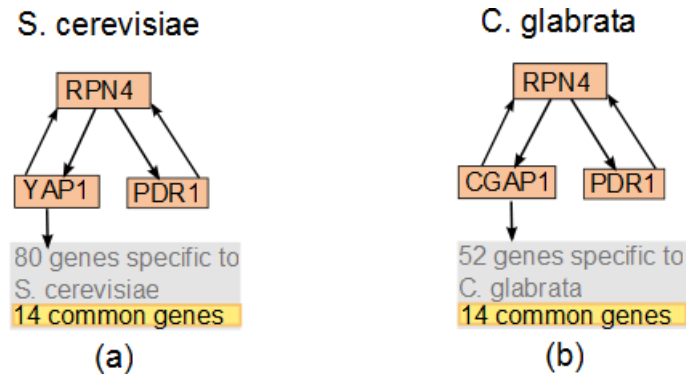


Figure 4.13: Conserved regulatory feedback module in the drug resistance pathway of *S. cerevisiae* and *C. glabrata*. Conserved genes are shown in colour and the genes which are not conserved in both species or not regulated by the same genes in both species are shown in grey.

Salin et al. (2008)). YAP1 is found to regulate around 80 different genes in *S. cerevisiae*, whereas, CGAP1, the *C. glabrata* homologue of *S. cerevisiae* YAP1 regulate only 52 genes. Among the genes regulated by YAP1 and CGAP1 only 14 genes have homologues in both species (Salin et al. (2008)). This suggests extensive evolutionary rewiring in the genetic circuit surrounding the above module.

The above examples give strong evidence in support of our theory. However, in case of multiple coupled feedback loops only a selected few are found conserved in other species. For example, in the *S. cerevisiae* cell cycle module, the positive feedback between CLB5 and SBF is not conserved in *S. pombe*. Instead, CIG2 and MBF, the *S. pombe* orthologs of *S. cerevisiae* CLB5 and SBF, are engaged in a negative feedback mechanism (Fig. 4.11). However, the other regulatory feed-

backs associated with them are conserved. In a separate instance, the feedback between GAL80, GAL4 and GAL2 of *S. cerevisiae* may not be present in the galactose metabolism module of *K. lactis* (Fig. 4.12). This is so because RAG1, the *K. lactis* homologue of GAL2 is thought to act as glucose transporter rather than galactose sensor (Betina et al. (2001)). One plausible explanation behind such phenomena is that when multiple feedback mechanisms are coupled together, not all of them have biologically vital roles in the operation of the whole module. Rather, only few of them have dominant kinetic roles, whereas the rest provides functional modulation. In that case it is possible that the most essential feedback mechanisms remain conserved. To confirm the above hypothesis we examined regulatory modules with coupled feedback loops where such disparities of conservation are observed. The nitrogen catabolite repression and glucose repression modules are excellent examples. In addition, necessary data on these two modules are available for our analysis. Hence, in the next chapter we focus on the kinetic properties of these two modules and explain the disparities in conservation of their different parts.

## 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

Pathway	Genes	Reference
Metabolic	YCL040W, YFR053C, YGL253W, YNR012W, YGL026C, YHR144C, YDR127W, YPL231W, YOR388C, YER069W, YER061C, YKL182W, YML106W, YMR271C, YBL039C, YJR103W, YBR243C, YGL065C, YKL067W, YKL060C, YBR018C, YBL035C, YBR278W, YDL102W, YDR121W, YIL139C, YIR008C, YJR006W, YJR043C, YKL045W, YNL102W, YNL262W, YOR330C, YPL167C, YPR175W, YAL012W, YKL106W, YLR027C, YDR502C, YLR180W, YLR245C, YBR084W, YGR204W, YKR080W, YBR110W, YBR117C, YPR074C, YGR192C, YJL052W, YJR009C, YJR057W, YDL232W, YEL002C, YGL022W, YGL226C-A, YJL002C, YML019W, YMR149W, YOR085W, YOR103C, YIR032C, YKL211C, YOR196C, YJR130C, YLL058W, YML082W, YDR178W, YJL045W, YKL141W, YKL148C, YLL041C, YLR164W, YMR118C, YEL029C, YNR027W, YDL131W, YDL182W, YBL082C, YLR359W, YBR035C, YGR208W, YPR069C, YOL140W, YDR297W, YHR104W, YDR300C, YGR043C, YLR354C, YLR028C, YMR120C, YER073W, YMR169C, YMR170C, YOR374W, YPL061W, YOR236W, YDR111C, YLR089C, YIL145C, YNL130C, YJL167W, YPL069C, YOL052C, YPL117C, YPR183W, YCR073W-A, YGR248W, YHR163W, YNR034W, YCR012W, YDR408C, YDL066W, YLR174W, YNL009W, YGL234W, YHR190W, YML035C, YHR123W, YNL141W, YBR154C, YDL140C, YDL150W, YDR045C, YDR156W, YDR404C, YGL070C, YHR143W-A, YIL021W, YJL140W, YJL148W, YJR063W, YKL144C, YNL113W, YNL151C, YNL248C, YNR003C, YOL005C, YOR116C, YOR151C, YOR207C, YOR210W, YOR224C, YOR340C, YOR341W, YPR010C, YPR110C, YPR187W, YPR190C, YPR127W, YOR074C, YBR263W, YLR058C, YLR209C, YLR420W, YBL068W, YER099C, YHL011C, YKL181W, YOL061W, YJL088W, YER070W, YGR180C, YIL066C, YJL026W, YGL148W, YER015W, YIL009W, YMR246W, YOR317W, YJL101C, YJL200C, YLR304C, YGR155W, YHR025W, YDR454C, YPL111W, YBR196C, YMR250W, YGR061C, YDR062W, YMR296C, YDR007W, YBR208C, YBR029C, YCL009C, YMR108W, YBR070C, YGL047W, YCL064C, YER086W, YER091C, YJL130C, YJR109C, YOR303W, YLR438W, YPR035W, YAL038W, YOR347C, YOR128C, YCR005C, YNR001C, YPR001W, YHL003C, YKL008C, YAR015W, YKL184W, YKR031C, YEL046C, YEL021W, YHL012W, YKL035W, YMR300C, YHR018C, YDR441C, YML022W, YJR105W, YDR019C, YKR097W, YBR221C, YER178W, YFL018C, YNL071W, YPL017C, YBR249C, YDR035W, YGR019W, YJR016C, YDR158W, YDL078C, YKL085W, YOL126C, YER081W, YIL074C, YPL262W, YLR303W, YMR062C, YOL058W, YER065C, YPR006C, YBR218C, YGL062W, YOR184W, YDL246C, YJR159W, YDR050C, YGR244C, YOR142W, YAL054C, YLR153C, YDL021W, YKL152C, YKR043C, YGR170W, YNL169C, YNL220W, YER026C, YDL080C, YGR087C, YLR044C, YBR176W, YLR239C, YCR053W, YBR006W, YOR095C, YNL037C, YOR136W, YNL241C, YIR029W, YGR256W, YHR183W, YJL121C, YAL062W, YDL171C, YDL215C, YOR375C, YBR252W, YKL024C, YDR354W, YHR216W, YLR432W, YML056C, YLR377C, YER090W, YFR014C, YGR240C, YMR205C, YOL016C, YMR217W, YKL216W, YHR063C, YER052C, YOR323C, YNR016C, YMR189W, YNL277W, YJR139C, YIR031C, YNL117W, YGR254W, YHR174W, YMR323W, YOR393W, YOL049W	Kanehisa et al. (2008)

## 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

Secondary metabolite synthesis	YBR196C, YDR050C, YGR192C, YJL052W, YJR009C, YOR095C, YCR012W, YBR117C, YPR074C, YPL028W, YGR043C, YLR354C, YJL167W, YPL069C, YDL066W, YLR174W, YNL009W, YKL060C, YJL121C, YPL111W, YGR254W, YHR174W, YMR323W, YOR393W, YNL037C, YOR136W, YHR018C, YCL040W, YFR053C, YGL253W, YGL026C, YPL117C, YCR073W-A, YGR248W, YHR163W, YNR034W, YJL088W, YBR249C, YDR035W, YNL241C, YGR240C, YLR377C, YMR205C, YDL078C, YKL085W, YOL126C, YOL058W, YDL021W, YKL152C, YKR043C, YML126C, YGR256W, YHR183W, YBR019C, YKL106W, YLR027C, YPL262W	Kanehisa et al. (2008)
MAPK mediated stress response	YBR200W, YER118C, YBL105C, YHL007C, YDL235C, YLR362W, YJL095W, YLR113W, YNL271C, YHL007C, YHL007C, YJL128C, YGR040W, YAL041W, YLR229C, YNL098C, YLR362W, YLR362W, YLR342W, YOR231W, YGR032W, YML004C, YHR084W, YHR030C, YDL159W, YLR332W, YBR083W, YMR043W, YDL159W, YPL089C, YLR229C, YCL027W, YOR212W, YBL016W, YKL178C, YGR088W, YPR165W, YER118C, YFL026W, YHR005C, YHR084W, YNL053W, YJL157C, YJR086W, YIL147C, YLR006C, YDR461W, YNL145W, YMR037C, YKL062W, YLR182W, YER111C, YOR212W, YJR086W, YDR480W, YPL049C, YPL187W, YGL089C, YCR073C, YNR031C, YPL049C, YDR480W, YDL159W, YBL105C, YCL027W, YHL007C, YHR005C, YER118C, YNL271C, YDL235C, YBR200W, YOR231W, YHR084W, YLR229C, YHR030C, YLR229C, YGR088W, YLR362W, YBR083W, YJL095W, YGR040W, YIL147C, YLR362W, YLR362W, YPL089C, YML004C, YMR043W, YHL007C, YNL053W, YLR342W, YHL007C, YPR165W, YAL041W, YNL098C, YLR332W, YLR006C, YER118C, YLR113W, YKL178C, YFL026W, YJL157C, YJL128C, YBL016W, YDL159W, YHR084W, YGR032W	Kanehisa et al. (2008); Klipp et al. (2005); Kofahl and Klipp (2004); Paliwal et al. (2007); Ross et al. (2000.)
Nitrogen response, Glucose sensing, Galactose metabolism	YBR020w, YLR081w, YDR009w, YPL248c, YML051w, YDL194w, YDL138w, YHR094c, YMR011w, YDR345c, YHR092c, YOR047c, YDR277c, YGL035c, YGL209w, YDR477w, YIL162w, YER040w, YNL229c, YFL021w, YKR034w, YJL110C, YLR113w, YER075c, YNL167c, YOR208w, YIL046w, YNL103w, YML027w, YPR065w, YOL116W, YMR016c, YDL020c, YML007w, YGL013c,	Boczko et al. (2005); Kaniak et al. (2004); Ramsey et al. (2006)
Transcription regulation	YBR202W, YPL049C, YDR043C, YDR451C, YDR207C, YCL067C, YML027W, YER040W, YOR372C, YJR060W, YMR070W, YGL073W, YAL040C, YOL148C, YFR034C, YLR183C, YLR182W, YDR146C, YPR072W, YML007W, YJL110C, YPL016W, YKL112W, YDR138W, YPR065W, YPR104C, YBR083W, YDR259C, YLR451W, YKL109W, YMR043W, YKL062W, YKR034W, YDR123C, YGL209W, YGL207W, YPL248C, YBR112C, YGL162W, YJL176C, YMR280C, YJR147W, YIR018W, YNL068C, YNL103W, YLR399C, YJL168C, YCR065W, YFL021W, YER111C, YKL043W, YHR084W, YHL027W, YLR013W, YNL216W, YOR028C, YJR127C, YMR016C, YLR131C, YEL009C, YLR256W, YBR049C, YGL096W, YPL177C	Lee et al. (2002); Luscombe et al. (2004b)
Cell cycle	YAL040c, YMR199w, YPL256c, YPR120c, YGR109c, YLR079w, YGL003c, YLR079w, YGL003c, YGR108w, YPR119w, YGL116w, YDR146c, YFR028c, YDR113c, YMR043w	Chen et al. (2000)

## 4.4 Conservation of feedback mechanism in different signalling and cell cycle network of yeasts

Pathway	Genes	Reference
Protein ubiquitination	<p> YOR249C, YLR102C, YGL240W, YGL116W, YGL003C, YIR025W, YDR118W, YNL008C, YHR166C, YBL084C, YFR036W, YNL172W, YKL022C, YGR225W, YLR127C, YDL132W, YGR003W, YJL047C, YBR259W, YDR069C, YGL094C, YFR005C, YNL186W, YKR098C, YJL197W, YBL067C, YBR058C, YMR304W, YPL072W, YDL122W, YOR124C, YER151C, YER144C, YFR010W, YIL156W, YMR223W, YER098W, YHL013C, YFL044C, YPL020C, YIL031W, YHR134W, YPL096W, YDL117W, YER162C, YDR314C, YFR004W, YDL216C, YOR261C, YHR165C, YJR099W, YPL003W, YPR180W, YHR171W, YKL210W, YDR390C, YPR066W, YHR111W, YHR003C, YKL027W, YDR054C, YGL087C, YGR133W, YMR022W, YGL058W, YBR171W, YCL008C, YOR339C, YLR306W, YDR092W, YDR177W, YBR082C, YDR059C, YER100W, YEL012W, YDL064W, YNR007C, YBR165W, YDL008W, YDL074C, YLR394W, YLR323C, YHR115C, YNL116W, YJL157C, YDL013W, YOL013C, YOL133W, YLR427W, YER068W, YOR156C, YLR148W, YMR231W, YDR265W, YMR026C, YOL054W, YBR114W, YCR066W, YLR032W, YOR191W, YDR143C, YDR409W, YER116C, YDR103W, YDR460W, YKL034W, YGR184C, YLR024C, YAL002W, YBR062C, YDR128W, YDR266C, YHL010C, YKR017C, YLR247C, YMR247C, YOL138C, YPR093C, YDL190C, YKL059C, YIL030C, YEL019C, YLL036C, YML068W, YER125W, YDR457W, YJR036C, YGL141W, YKL010C, YDR313C, YLR128W, YDR255C, YMR119W, YIL046W, YML088W, YFL009W, YJR090C, YNL311C, YJL204C, YDR131C, YLR368W, YJL149W, YBR203W, YOR080W, YDR219C, YLR097C, YNL230C, YDR306C, YLR224W, YMR258C, YMR094W, YBR158W, YBR280C, YLR352W, YER143W, YGL181W, YOR138C, YKL090W, YOR042W, YDR273W, YML097C, YOL087C, YNR006W, YDL161W, YDR464W, YLR206W, YHL002W, YDR082W, YNL059C, YKL213C, YHR079C, YNL155W, YOR052C, YNR051C, YGR048W, YNL119W, YGL211W, YGR200C, YMR312W, YDL126C, YMR100W, YOR057W, YBL047C, YJR052W, YMR275C, YML111W, YNR069C, YOR197W, YDR260C, YMR316W, YBL057C, YPL002C, YIR011C, YDR002W, YNL159C, YJR102C, YPL084W, YNR068C, YML101C, YBR111W-A, YPL065W, YLR417W, YGL017W, YJR062C, YLR207W, YBR201W, YKL054C, YDR411C, YDR057W, YGL110C, YMR264W, YDR328C, YPL046C, YIL001W, YLR108C, YDR132C, YOR043W, YBL041W, YER012W, YPR103W, YJL001W, YFR050C, YMR314W, YOL038W, YGR135W, YML092C, YOR362C, YOR157C, YGR253C, YER094C, YGL011C, YHR200W, YER021W, YDL097C, YDR427W, YPR108W, YMR309C, YDL147W, YBR079C, YKL145W, YOR259C, YGL048C, YOR117W, YDR394W, YDL007W, YHL030W, YHR027C, YFL007W, YDR363W-A, YIL075C, YFR052W, YDL020C, YBR173C, YLR421C, YGR232W, YGL004C, YLR021W, YPL144W, YBR135W, YIL007C, YKL206C, YMR191W, YLR199C, YIL071C, YJR084W, YOL117W, YDR179C, YMR025W, YPR049C, YMR276W, YNR032C-A, YER007W, YEL037C, YIL148W, YKR094C, YLR167W, YDR139C, YDR510W, YLL039C, YML029W, YOL111C, YML013W, YBL058W, YDL091C, YMR067C, YDR330W, YJL048C, YBR273C, YIL008W, YPL149W, YBR170C, YBR217W </p>	Venancio et al. (2009)

## Chapter 5

# A comparative study of transcriptional feedback modules in different yeast species

The roles of biochemical feedback mechanisms in the evolution of organisms only begins to emerge. Many studies suggest that some "kernel" modules are evolutionary well preserved while their input and output often have adopted different functionalities ([Hinman et al. \(2009\)](#)). In our integrated networks we find many different kinds of feedback loops, e.g., loops made of TF-DNA interactions, protein-protein interactions, and composite loops made of both type of interactions. The feedback loops found in TF-DNA interactions are typically small, spanning two to five genes, whereas many long feedback mechanisms are found in PINs (see table [3.2](#)). However, in most cases, feedback loops are clus-



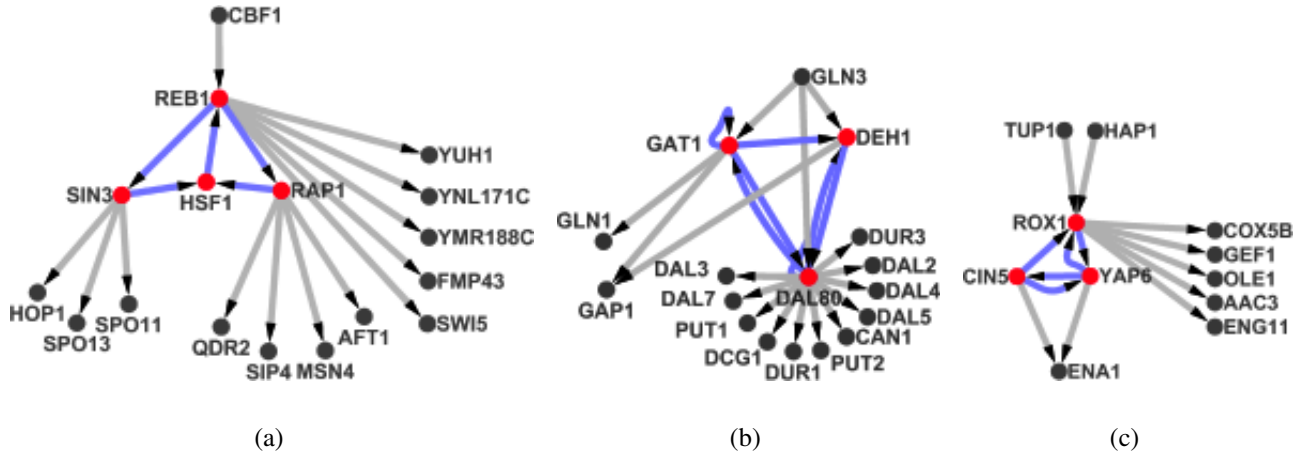


Figure 5.1: Cyclic modules reside in the hierarchical middle layer of yeast GRNs. The module shown in panel (a) is part of the yeast sporulation GRN. The modules shown in panel (b) and (c) are part of stress response GRNs. The cyclic modules are drawn in red nodes and blue edges. The subnetworks are generated from regulatory interaction data [Lee et al. \(2002\)](#); [Luscombe et al. \(2004b\)](#); [Yu and Gerstein \(2006\)](#). The regulatory hierarchy is made by imposing a partial order on the set of genes which form these subnetworks. We used only TF-DNA interaction data to build these subnetworks in order to simplify the visualization procedure.

tered together by means of coupling and nesting to form cyclic modules. Among them, those made of transcription regulation interactions preferentially lie in the middle level of regulatory hierarchy as shown in figure 5. In GRNs the middle layer of the regulatory hierarchy has a pivotal role in genetic decision making ([Yu and Gerstein \(2006\)](#)). Due to bifurcative behaviour of feedback mechanisms in response to parametric perturbations (see below) such phenomena alter their functional behaviour profoundly and thereby affect the decision making procedures of GRNs. Thus, evolutionary adaptation is more likely to be successful outside such feedback modules, in the acyclic parts.

---

We observe that feedback loops reside in a clustered manner both in transcription regulation and protein interaction networks. Such clusters occur due to coupling and nesting of several feedback mechanisms with each other. Thus, a key question is whether all the feedback loops in these modules are similarly important to their operations, or, some of them are biologically vital and the rest provide functional assistance. From here on we shall call the biologically vital loops as 'dominant' feedback loop and the ones which provide functional assistance as 'auxiliary' loop. The best way to find out whether a feedback loop is 'dominant' or 'auxiliary' is to determine phenotypes after disrupting different parts of these feedback loops. If disrupting a feedback loop by knocking out one or more of its component produces deleterious phenotype then it can be thought of as a 'dominant' feedback and if such knockout experiments do not produce any harmful effect then it can be thought of as an 'auxiliary' loop. Such biological experiments were performed by other scientists in the past on certain *S. cerevisiae* pathways e.g. nitrogen catabolite repression pathway [Boczko et al. \(2005\)](#), carbon catabolite repression circuit [Kaniak et al. \(2004\)](#), pheromone response pathway [Paliwal et al. \(2007\)](#), osmoregulation pathway [Klipp et al. \(2005\)](#) etc. Though these studies determined which feedback loops in those pathways are vital, the role of the auxiliary feedback loops in these pathways were not elucidated. Hence, it is interesting to find out the roles of the auxiliary feedback mechanisms of these pathways and investigate whether the dominant and auxiliary feedback loops of these pathways evolve differently.

In a study, [Kwon and Cho \(2008\)](#) found that feedback loops are generally clustered in signalling pathways in a coherent manner, i.e.

## **5.1 Dominant and auxiliary feedback in transcriptional feedback modules**

---

same type of feedback loops (either positive or negative) are coupled together to form these clusters. **Kwon and Cho (2008)** also proposed that coherent couplings can enhance the robustness of a signalling network and tend to be evolutionarily stable. In this chapter, we investigate those cases where at least one of the coupled feedback is not found to be evolutionarily conserved in other species to find the reasons behind differential evolution in different parts of these feedback clusters.

### **5.1 Dominant and auxiliary feedback in transcriptional feedback modules**

To understand the functional roles of individual feedback loops in a transcriptional feedback module we investigate the dynamic properties of two of the most well studied transcriptional feedback modules found in yeast, i.e. nitrogen catabolite repression circuit (NCR) and carbon catabolite repression circuit (CCR). The detailed analysis of these modules are as follows:

#### **5.1.1 Nitrogen catabolite repression (NCR) circuit**

Nitrogen is a basic building block of biomolecules such as proteins, nucleic acids etc. All living cells need to induct nitrogen from growth

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

---

medium for many biological processes. Nitrogen catabolite repression is the physiological process by which budding yeast prefers good nitrogen (glutamine, asparagine, ammonia) sources over the poor ones (allantoin, proline, urea). A schematic diagram of the NCR circuit is shown in figure 5.1.1 (Boczko et al. (2005)).

The yeast NCR circuit gets activated in the presence of limiting or poor nitrogen supply and transcribes the genes which transport and degrade poor nitrogen sources. On the other hand, in the presence of excess nitrogen NCR circuit is repressed and the repression of genes which take part in transportation and degradation of good nitrogen sources is released. In the NCR circuit, nitrogen scarcity is sensed by Ure2p and Gln3p proteins which then regulate the transcription of GAT1, DAL80 and DEH1 genes. GAT1, DAL80 and DEH1 genes regulate each other's transcriptions and give rise to a module made of two coupled feedback loops, one made of GAT1 and DAL80 and the other made of DAL80 and DEH1. In this section our objective is to analyse the dynamics of NCR circuit in order to determine which of its feedback loops has 'dominant' and which one has 'auxiliary' functional role. Our intuition is that the 'dominant' feedback loop among these two should be found conserved in other yeast species as well. Following Boczko et al. (2005) we develop a mathematical model to simulate the functioning of the NCR circuit.

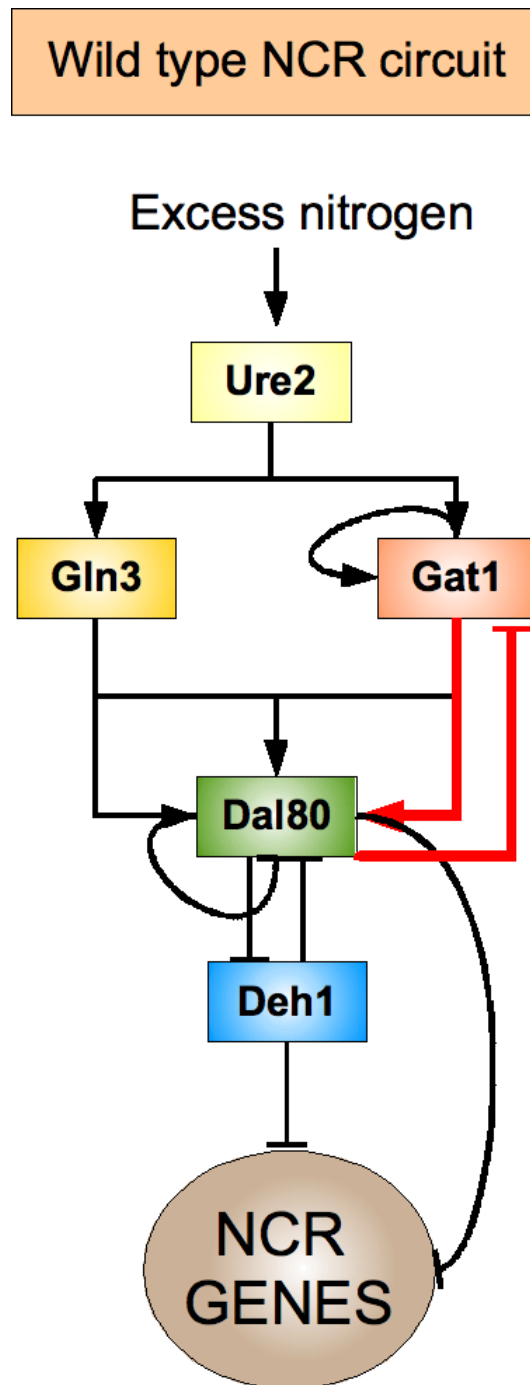


Figure 5.2: Wild type nitrogen catabolite repression circuit of budding yeast. Arrows indicate activation and blunt arrows indicate repression. The feedback module in this network is formed due to transcription regulations of GAT1, DAL80 and DEH1 genes by each other.

### 5.1.2 Dynamics of wild type NCR circuit

A detailed mathematical representation of the NCR switch can be found in Boczko et al. (2005). The kinetic model described by Boczko et al. (2005) includes the time delays between transcription and translation( $\delta_i$ ), the time delays during shuttling of proteins in and out of nucleus( $\tau$ ) and changes in protein concentrations due to cell cycle. We simplified the model assuming that the changes in protein concentrations due to cell cycle are slower compared to the changes due to nitrogen scarcity in the medium, thus assuming such rates to be constants. Following Boczko et al. (2005), we also ignored the time delays of the model in our analysis as a further simplification. The functions representing the binding of transcription factors with the DNA molecules are also changed in our model. We used Michaelis Menten kinetics to model the binding of Transcription factors with their corresponding DNA molecules. Most of the parameters are unaltered in the new model. However, the basal production rate of both URE2 ( $r_u$ ) and GLN3 ( $r_g$ ) mRNA molecules, the binding affinity of GAT1 ( $a_1$ ), DAL80 ( $a_2$ ) and DEH1 ( $a_3$ ) and the Michaelis constants  $k_{ij}$  were subjected to further parameter inference using simulated annealing based inference technique (Schmidt and Jirstrand (2006)). The Hill Coefficients are set to 2 following common convention. The resulting model is shown in equation 5.1 and its parameter values are shown in table 5.1

### 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

---

$$\begin{aligned}
\frac{dx(t)}{dt} &= \mathcal{K}_{imp}(\xi(t - \tau)) - \mathcal{K}_{exp}(x(t)) - \alpha(N)x(t) \\
\frac{dX(t)}{dt} &= r_g - \beta(N)X(t) \\
\frac{d\xi(t)}{dt} &= T(X(t - \delta_1)) - \gamma(N)\xi(t) - k_f(N)\xi(t)\mu(t) + k_r(N)C(t) \\
&\quad + \mathcal{K}_{exp}(x(t - \tau)) - \mathcal{K}_{imp}(\xi(t)) \\
\frac{dU(t)}{dt} &= r_u - \theta(N)U(t) \\
\frac{d\mu(t)}{dt} &= T(U(t - \delta_2)) - \kappa(N)\mu(t) - k_f(N)\xi(t)\mu(t) + k_r(N)C(t) \\
\frac{dC(t)}{dt} &= k_f(N)\xi(t)\mu(t) - k_r(N)C(t) \\
\frac{dy(t)}{dt} &= \mathcal{K}_{imp}(\psi(t - \tau)) - \mathcal{K}_{exp}(y(t)) - \pi(N)y(t) \\
\frac{dY(t)}{dt} &= \mathcal{H}_1(x, y, z, w) - \phi(N)Y(t) \\
\frac{d\psi(t)}{dt} &= T(Y(t - \delta_3)) - \eta(N)\psi(t) + \mathcal{K}_{exp}(y(t - \tau)) - \mathcal{K}_{imp}(\psi(t)) \\
\frac{dz(t)}{dt} &= \mathcal{K}_{imp}(\zeta(t - \tau)) - \mathcal{K}_{exp}(z(t)) - \rho(N)z(t) \\
\frac{dZ(t)}{dt} &= \mathcal{H}_2(x, y, z) - \nu(N)Z(t) \\
\frac{d\zeta(t)}{dt} &= T(Z(t - \delta_4)) - \chi(N)\zeta(t) + \mathcal{K}_{exp}(z(t - \tau)) - \mathcal{K}_{imp}(\zeta(t)) \\
\frac{dw(t)}{dt} &= \mathcal{K}_{imp}(w(t - \tau)) - \mathcal{K}_{exp}(w(t)) - \varphi(N)w(t) \\
\frac{dW(t)}{dt} &= \mathcal{H}_3(x, y, z) - \varrho(N)W(t) \\
\frac{d\omega(t)}{dt} &= T(W(t - \delta_5)) - \varpi(N)\omega(t) + \mathcal{K}_{exp}(w(t - \tau)) - \\
&\quad \mathcal{K}_{imp}(\omega(t))
\end{aligned} \tag{5.1}$$

The definition of the conjugate variables in 5.1 are as follows.

$$\begin{aligned}
\mathcal{K}_{imp}(x) &= \mathcal{K}_{exp}(x) = \frac{A_k x(1 + B_k x)}{C_k + x} \\
T(x) &= \frac{A_T x^2}{B_T + x} \\
\mathcal{H}_1(x, y, z, w) &= a_1 \frac{\left(\frac{x}{k_{11}}\right)^2 + \left(\frac{z}{k_{12}}\right)^2}{1 + \left(\frac{x}{k_{11}}\right)^2 + \left(\frac{z}{k_{12}}\right)^2 + \left(\frac{y}{k_{13}}\right)^2 + \left(\frac{w}{k_{14}}\right)^2} \\
\mathcal{H}_2(x, y, z) &= a_2 \frac{\left(\frac{x}{k_{21}}\right)^2 + \left(\frac{z}{k_{22}}\right)^2}{1 + \left(\frac{x}{k_{21}}\right)^2 + \left(\frac{z}{k_{22}}\right)^2 + \left(\frac{y}{k_{23}}\right)^2} \\
\mathcal{H}_3(x, y, z) &= a_3 \frac{\left(\frac{x}{k_{31}}\right)^2 + \left(\frac{z}{k_{32}}\right)^2}{1 + \left(\frac{x}{k_{31}}\right)^2 + \left(\frac{z}{k_{32}}\right)^2 + \left(\frac{y}{k_{33}}\right)^2}
\end{aligned}$$

In equation 5.1,  $x(t)$ ,  $y(t)$ ,  $z(t)$ ,  $w(t)$  represents the concentrations of nuclear Gln3, Dal80, Gat1 and Deh1 proteins.  $\xi(t)$ ,  $\mu(t)$ ,  $\psi(t)$ ,  $\zeta(t)$  and  $\omega(t)$  represent the concentrations of the cytoplasmic Gln3, Ure2, Dal80, Gat1, Deh1 proteins.  $X(t)$ ,  $U(t)$ ,  $Y(t)$ ,  $Z(t)$ ,  $W(t)$  represent the concentrations of mRNA molecules produced by GLN3, URE2, DAL80, GAT1, DEH1.  $T(x)$  simulates the translation initiation of mRNA molecules in the cytoplasm (Boczko et al. (2005)).  $\mathcal{K}_{imp}$  and  $\mathcal{K}_{exp}$  simulates the shuttling of proteins in and out of nucleus.  $\alpha(N)$ ,  $\beta(N)$ ,  $\gamma(N)$ ,  $\theta(N)$ ,  $\kappa(N)$ ,  $\pi(N)$ ,  $\phi(N)$ ,  $\eta(N)$ ,  $\rho(N)$ ,  $\nu(N)$ ,  $\chi(N)$ ,  $\varphi(N)$ ,  $\varrho(N)$  and  $\varpi(N)$  are degradation rates which, in this case, depend on nitrogen concentration in the medium (Boczko et al. (2005)). In our analysis, we kept the degradation parameters constants and studied the behaviour of the system for different values of these parameters which, in effect, reflects the fluctuation in nitrogen concentrations. Parameter values for equation 5.1 are shown in table 5.1 . We executed multiple runs of simulated



### 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

Params	Values	Params	Values	Params	Values	Params	Values
$A_k$	60	$\kappa$	0.009	$\varpi$	0.009	$k_{14}$	0.198
$B_k$	0.03	$\pi$	0.00485	$k_f$	1e-05	$a_2$	0.108
$C_k$	4	$\phi$	0.095	$k_r$	0.13	$k_{21}$	0.032
$A_T$	0.1	$\eta$	0.009	$r_g$	0.0031	$k_{22}$	0.028
$B_T$	0.26	$\rho$	3.6	$r_u$	0.00063	$k_{23}$	0.016
$\alpha$	0.00485	$\nu$	0.009	$a_1$	0.181	$a_3$	0.0124
$\beta$	0.022	$\chi$	0.067	$k_{11}$	0.00055	$k_{31}$	0.1
$\gamma$	0.009	$\varphi$	0.00485	$k_{12}$	0.162	$k_{32}$	0.2
$\theta$	0.022	$\varrho$	0.022	$k_{13}$	0.000852	$k_{33}$	0.186

Table 5.1: Parameter values of equation 5.1 which represents the mathematical model of Nitrogen Catrabolite Repression Circuit of wild type yeast cells.

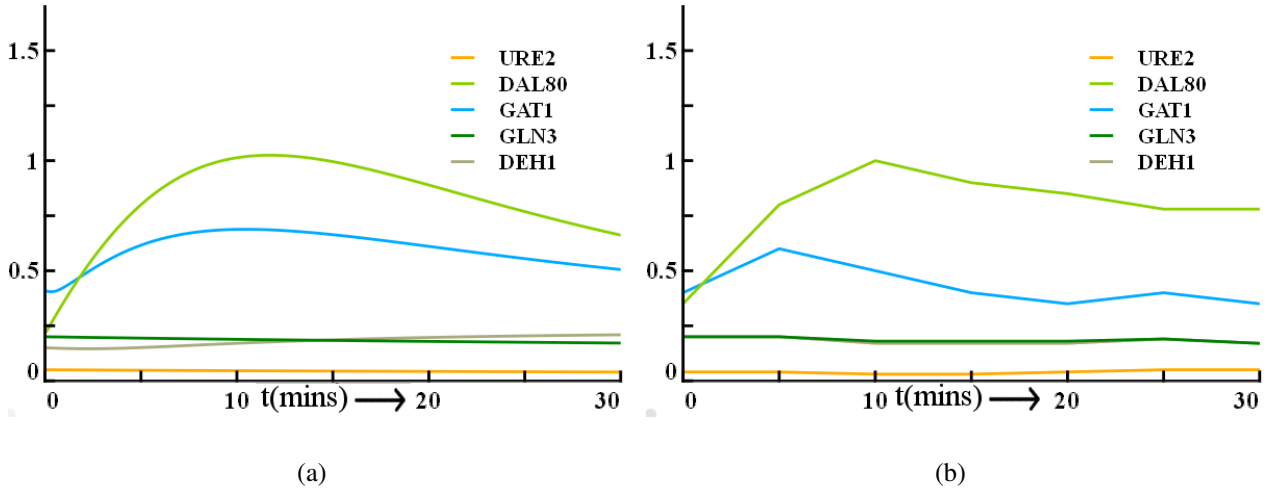


Figure 5.3: Simulation of the model shown in equation 5.1 (a) and the observed expression levels (b) of components of the NCR circuit during nitrogen starvation.

annealing on our model for [Boczko et al. \(2005\)](#)'s data set and collected the best fit among them. The parameter values are given in table 5.1.

In figure 5.3 we show the result of simulation of the model shown in equation 5.1 with the parameter values shown in table 5.1 and the expression levels observed by [Boczko et al. \(2005\)](#).

### 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

---

We characterise the autonomous behaviour of the wild type NCR circuit by carrying out bifurcation analysis of the mathematical model shown in equation 5.1. Bifurcation analysis reveals that the model of wild type cells can exhibit bistability. Bistability provides twofold advantages to the wild type NCR circuit. Firstly, it introduces a switch like behaviour via DEH1 gene which flips between 'on' and 'off' state due to fluctuations in kinetic rate constants, especially the degradation constants (figure 5.4). The degradation constants depend on the nitrogen concentration in the growth medium Boczko et al. (2005). Hence, in effect, the system switches the DEH1 gene 'on' and 'off' due to fluctuations in the nitrogen concentration, thus gains extra control over NCR genes. Secondly, it realises a cellular memory system where DEH1 keeps a back up of the initial state of DAL80, which is helpful to modulate its behaviour during future episodes of nitrogen scarcity. The responses of the wild type NCR model to different initial concentrations of DAL80 and DEH1 mRNA are shown in figure 5.5. A truth table of the responses for different combinations of high ( $[X] \geq 0.5$ ) and low ( $[X] < 0.5$ ) initial concentrations of DAL80 and DEH1 genes are shown in table 5.2. Table 5.2 shows that the wild type NCR circuit resembles a 'D' type flip flop which has the boolean logic  $[DEH1]^{n+1} = \overline{[DAL80]^n}$ .

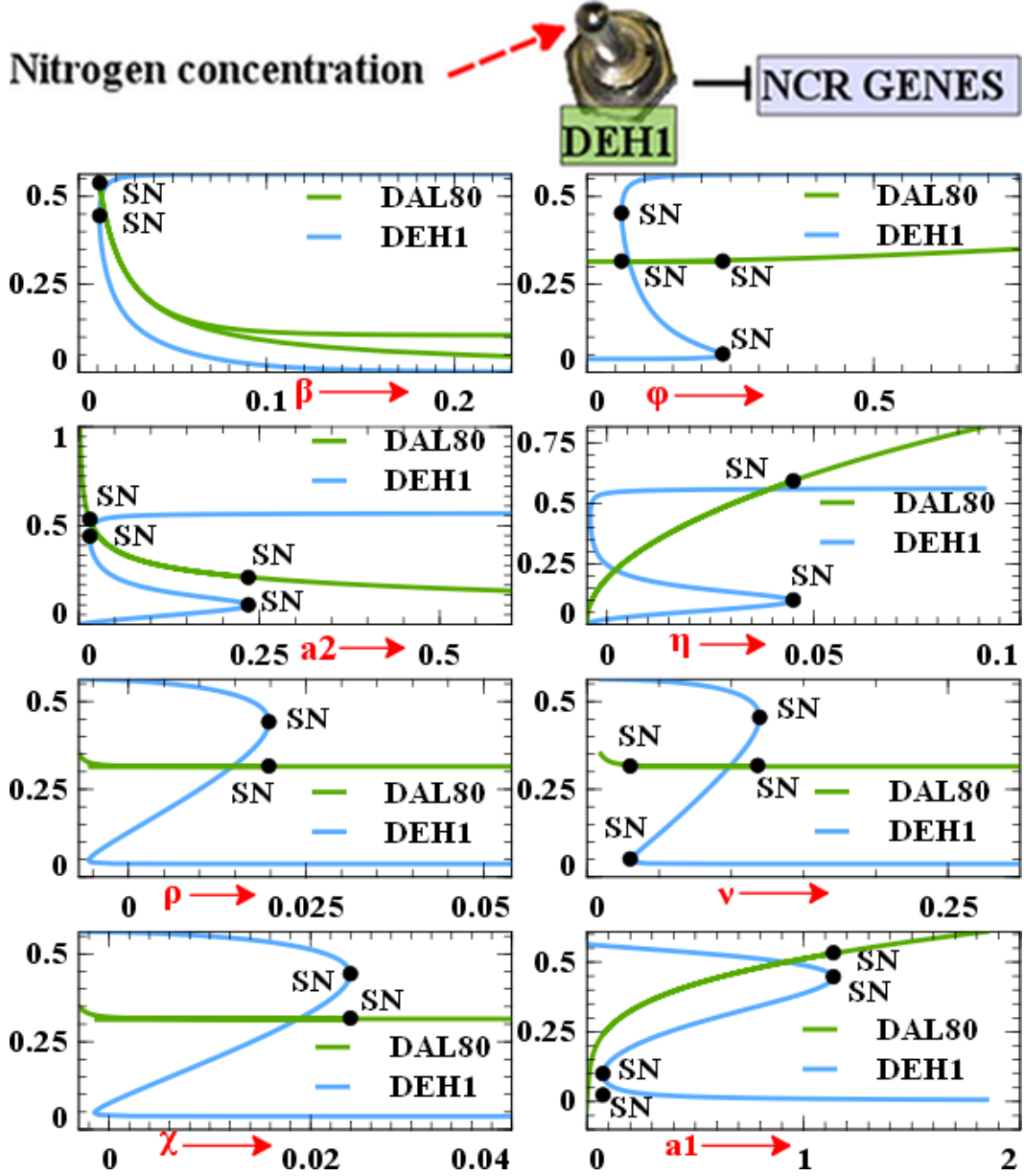


Figure 5.4: Bifurcation diagrams for the wild type NCR model shown in equation 5.1. The control mechanism of the NCR circuit relies on nitrogen state dependent control of its degradation rates (Boczko et al. (2005)). The observed saddle node bifurcations (SN) are mainly due to fluctuations in degradation rates ( $\beta$ ,  $\phi$ ,  $\eta$ ,  $\rho$ ,  $\mu$ ,  $\chi$ , see equation 5.1). Interestingly, the concentration changes of DAL80 caused by bifurcation are negligible, whereas DEH1 goes through a stiff hysteresis curve which switches its state between 'on' ( $> 0.5$ ) and 'off' ( $< 0.5$ ). Hence, in the wild type NCR circuit DEH1 acts as a nitrogen state dependent toggle switch.

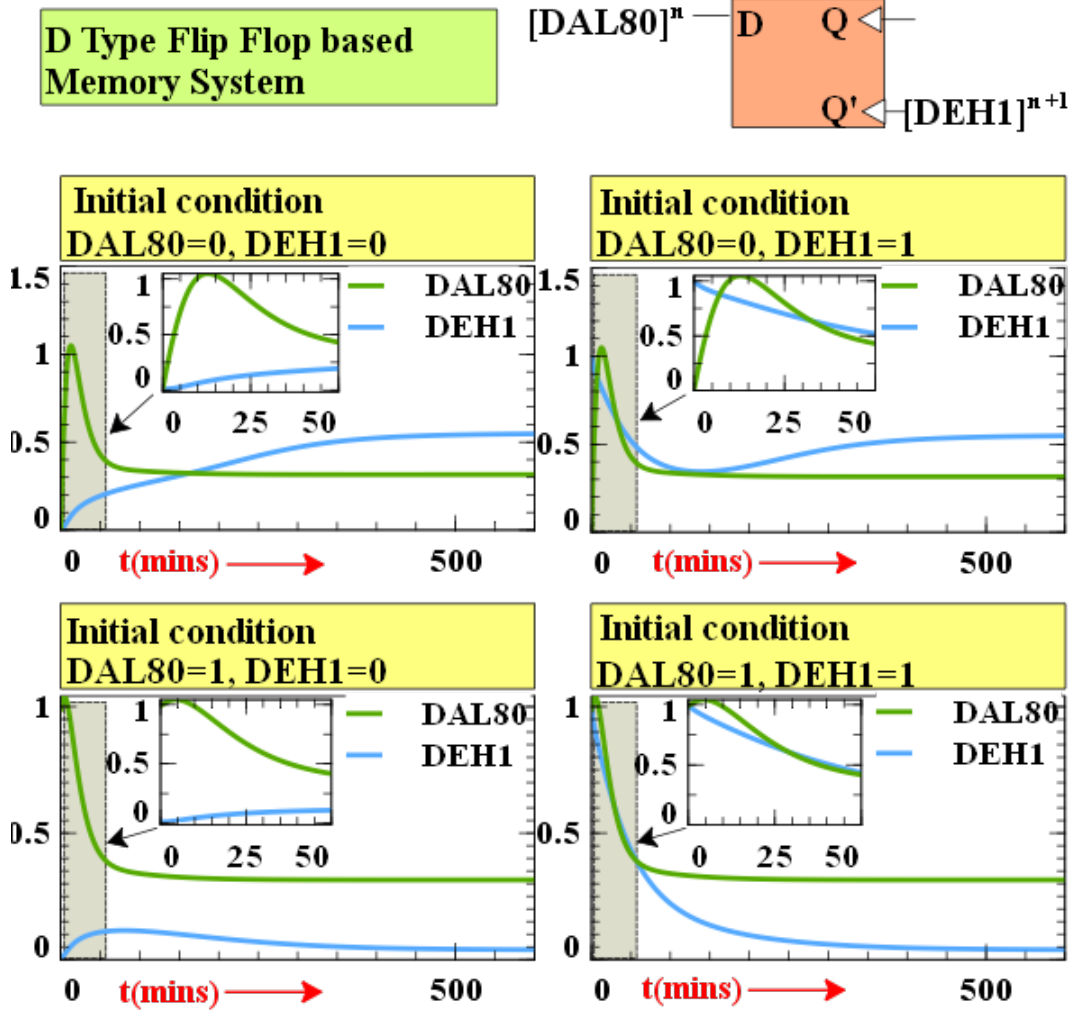


Figure 5.5: Responses of the wild type NCR circuit due to changes in initial concentrations of DAL80 and DEH1 mRNAs. We have simulated the responses of the wild type NCR circuit for four different combinations of initial concentrations, i.e. ( $[DAL80] = 0$ ,  $[DEH1] = 0$ ), ( $[DAL80] = 0$ ,  $[DEH1] = 1$ ), ( $[DAL80] = 1$ ,  $[DEH1] = 0$ ) and ( $[DAL80] = 1$ ,  $[DEH1] = 1$ ). It can be seen from this figure that in steady state DAL80 always comes down to 'off' state. However, the steady state response of DEH1 gene depends on the initial state of DAL80 and follows the Boolean logic  $[DEH1]^{n+1} = \overline{[DAL80]}^n$  which realizes a D type flip flop (truth table shown in table 5.2). The importance of such a backup system lies in the fact that the functioning of DAL80 is mainly transient and always comes down to its 'off' state needing a memory to remember its initial condition for future reference.

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

$[Dal80]^n$	$[Deh1]^n$	$[Dal80]^{n+1}$	$[Deh1]^{n+1}$
0	0	0	1
0	1	0	1
1	0	0	0
1	1	0	0

Table 5.2: Truth table for the cellular memory system in the wild type NCR circuit.  $[DAL80]^n$  and  $[DEH1]^n$  represent the current state of DAL80 and DEH1 gene and  $[DAL80]^{n+1}$  and  $[DEH1]^{n+1}$  represent their next states. '1' and '0' represents high and low concentrations. High concentration of mRNA occurs when a gene is upregulated and vice versa.

### 5.1.3 DEH1 as an auxiliary gene

To understand the functional role of the genes involved in yeast NCR circuit [Coffman et al. \(1997\)](#) performed a knock out analysis. [Coffman et al. \(1997\)](#) found that knocking out GAT1 and DAL80 produces lethal phenotype in nitrogen scarce medium, whereas a  $deh1\Delta$  mutant does not produce any deleterious phenotypes. This finding suggests that GAT1 and DAL80 are essential to the functioning of the NCR circuit, whereas the DEH1 gene provides non essential functional benefits. On the other hand, an wild type NCR circuit has a feedback module which consists of two coupled small transcriptional feedback mechanisms whereas a  $deh1\Delta$  mutant has only one feedback mechanism made of GAT1 and DAL80 and lacks the extra feedback loop which appears due to the presence of the DEH1 gene (see figure 5.1.3). To investigate the roles of the feedback loop made of DEH1 we analysed the dynamic behaviour of the NCR circuit of a  $deh1\Delta$  mutant and compared it with that of a wild type NCR circuit.

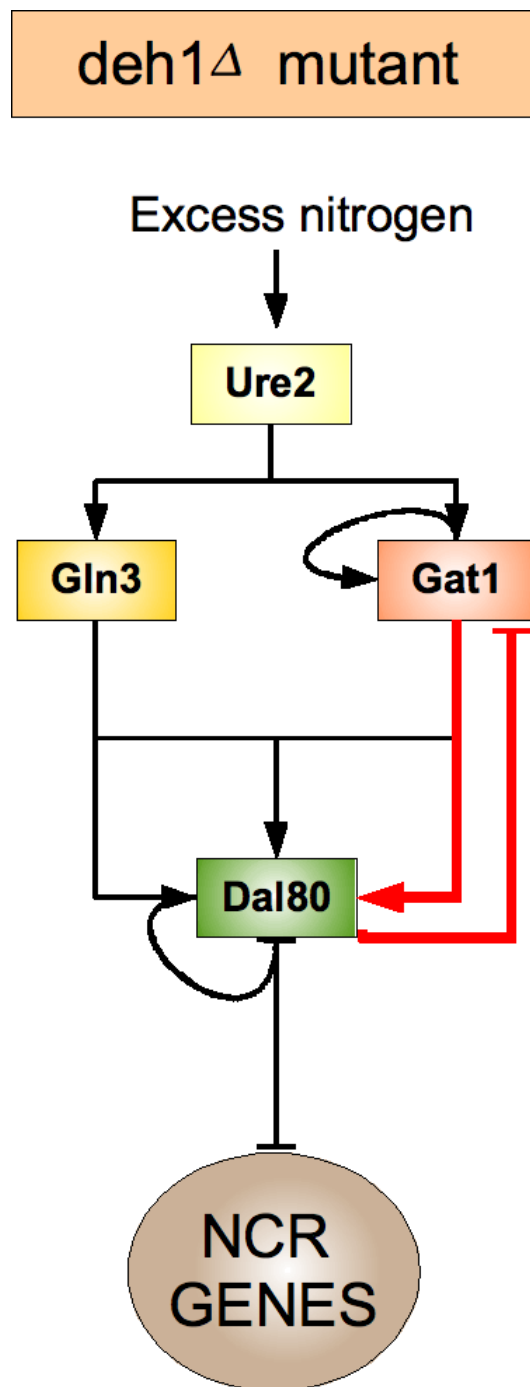


Figure 5.6: Schematic diagram of a NCR circuit in a *deh1Δ* mutant.

### 5.1.4 Dynamics of a $\text{deh1}\Delta$ mutant

The ODE model for the  $\text{deh1}\Delta$  mutant is shown in equation 5.2. The relevant parameter values are unchanged and can be found in table 5.1.

$$\begin{aligned}
\frac{dx(t)}{dt} &= \mathcal{K}_{imp}(\xi(t - \tau)) - \mathcal{K}_{exp}(x(t)) - \alpha(N)x(t) \\
\frac{dX(t)}{dt} &= r_g - \beta(N)X(t) \\
\frac{d\xi(t)}{dt} &= T(X(t - \delta_1)) - \gamma(N)\xi(t) - k_f(N)\xi(t)\mu(t) + k_r(N)C(t) \\
&\quad + \mathcal{K}_{exp}(x(t - \tau)) - \mathcal{K}_{imp}(\xi(t)) \\
\frac{dU(t)}{dt} &= r_u - \theta(N)U(t) \\
\frac{d\mu(t)}{dt} &= T(U(t - \delta_2)) - \kappa(N)\mu(t) - k_f(N)\xi(t)\mu(t) + k_r(N)C(t) \\
\frac{dC(t)}{dt} &= k_f(N)\xi(t)\mu(t) - k_r(N)C(t) \\
\frac{dy(t)}{dt} &= \mathcal{K}_{imp}(\psi(t - \tau)) - \mathcal{K}_{exp}(y(t)) - \pi(N)y(t) \\
\frac{dY(t)}{dt} &= \mathcal{H}_1^\Delta(x, y, z, w) - \phi(N)Y(t) \\
\frac{d\psi(t)}{dt} &= T(Y(t - \delta_3)) - \eta(N)\psi(t) + \mathcal{K}_{exp}(y(t - \tau)) - \mathcal{K}_{imp}(\psi(t)) \\
\frac{dz(t)}{dt} &= \mathcal{K}_{imp}(\zeta(t - \tau)) - \mathcal{K}_{exp}(z(t)) - \rho(N)z(t) \\
\frac{dZ(t)}{dt} &= \mathcal{H}_2^\Delta(x, y, z) - \nu(N)Z(t) \\
\frac{d\zeta(t)}{dt} &= T(Z(t - \delta_4)) - \chi(N)\zeta(t) + \mathcal{K}_{exp}(z(t - \tau)) - \mathcal{K}_{imp}(\zeta(t))
\end{aligned} \tag{5.2}$$

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

---

Where the definitions for the conjugate variables are as follows.

$$\begin{aligned}
 \mathcal{K}_{imp}(x) &= \mathcal{K}_{exp}(x) = \frac{A_k x (1 + B_k x)}{C_k + x} \\
 T(x) &= \frac{A_T x^2}{B_T + x} \\
 \mathcal{H}_1^\Delta(x, y, z) &= a_1 \frac{\left(\frac{x}{k_{11}}\right)^2 + \left(\frac{z}{k_{12}}\right)^2}{1 + \left(\frac{x}{k_{11}}\right)^2 + \left(\frac{z}{k_{12}}\right)^2 + \left(\frac{y}{k_{13}}\right)^2} \\
 \mathcal{H}_2^\Delta(x, y, z) &= a_2 \frac{\left(\frac{x}{k_{21}}\right)^2 + \left(\frac{z}{k_{22}}\right)^2}{1 + \left(\frac{x}{k_{21}}\right)^2 + \left(\frac{z}{k_{22}}\right)^2 + \left(\frac{y}{k_{23}}\right)^2}
 \end{aligned} \tag{5.3}$$

Bifurcation analysis of the  $deh1\Delta$  NCR model shown in equation 5.2 reveals that it does not exhibit bistability and hence lacks the memory module and the extra nitrogen controlled switching mechanism found in wild type cells. The bifurcation diagrams of the  $deh1\Delta$  model are shown in figure 5.7.

### 5.1.5 Dominant and auxiliary feedback in the budding yeast NCR circuit

The dynamic analysis of both the wild type and  $deh1\Delta$  mutant NCR model suggests that the DEH1 gene provides the wild type NCR circuit with both a switch like capability and a cellular memory system. However, since a  $deh1\Delta$  mutant does not produce any deleterious phenotype



## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

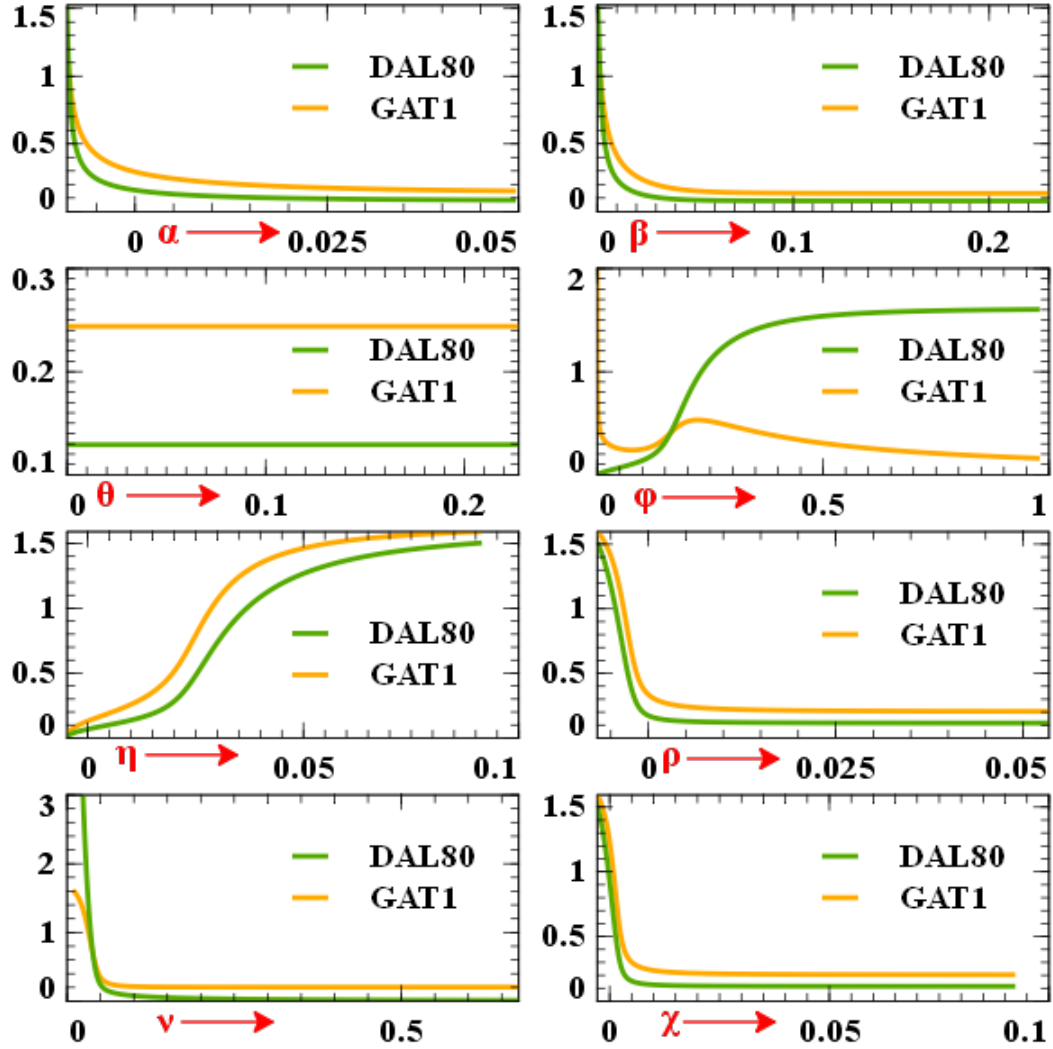


Figure 5.7: Bifurcation diagrams of the *deh1* $\Delta$  mutant NCR model. No hysteresis occurs due to perturbations in the reaction rates of the *deh1* $\Delta$  mutant NCR model. Hence, it does not produce bistability. For the convenience of visualisation we have shown the homeostasis levels GAT1 and DAL80 only.

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

---

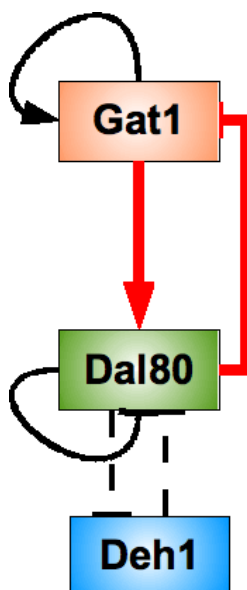


Figure 5.8: The 'dominant' and 'auxiliary' feedback loops of budding yeast NCR circuit. The thick red arrows indicate the interactions that forms the main feedback mechanism of budding yeast NCR circuit. The auxiliary feedback loops are represented by black dotted lines.

we conclude that the DEH1 mediated switching and memory module is not essential to the operation of the NCR module but an auxiliary upgrade to it. On the other hand, since mutating GAT1 or DAL80 produces lethal phenotype in nitrogen scarce medium we conclude that these two genes are essential in the operation of the NCR circuit. Hence, we suggest that the feedback mechanism made of DAL80 and GAT1 is biologically most vital to its operation and can be termed as the 'dominant' feedback whereas the feedback mediated by DEH1 is beneficial but not essential and can be termed as 'auxiliary' feedback(see figure 5.8). Due to the essential biological role of the main feedback we predict that it should be conserved in many different yeast species. To validate our prediction we carry out a phylogenetic analysis on the yeast NCR genes.

### 5.1.6 Conservation of the main feedback loop of budding yeast NCR module

In NCR circuit, Gat1, Dal80 and Deh1(Gzf3) are all GATA binding Zn finger protein encoding genes(GZF), however, Dal80 and Gzf3 are homologous to each other and they both have an extra Leucine zipper(LZ) subunit which makes them negative regulators of NCR genes. In almost all the yeast species of Saccharomycetaceae family, at least one GZF and one GZF-LZ can be found which are homologous to *S. cerevisiae* Gat1 and Dal80(information collected from KEGG homology database, [Rossa and Peter \(2005\)](#), and [Low and Atchly \(2000\)](#)). However, it is yet to be confirmed whether these homologs engage in any feedback interaction or not. In case of more distant yeast species such as *S. pombe* and *N. crassa*, the GZF encoding genes do not have such clear homology with their *S. cerevisiae* counterparts. It is argued by previous researchers that the ancestral yeast had only one GZF encoding protein ([Rossa and Peter \(2005\)](#)). The genes with double GZF-GZF encoding subunits found in species like *S. pombe* might have originated from whole genome duplication events and the ones with an extra LZ subunit evolved much later ([Rossa and Peter \(2005\)](#)), most probably during the divergence of Saccharomycetaceae family. In other words, the homologs of Gat1 and Dal80 are mainly found in Saccharomycetaceae family and not other yeast species because the ancestral yeast did not have both of them, they originated much later in the yeast lineage. A graphical representation of the evolution of GATA binding protein encoding genes is shown in figure [5.9](#).

## **5.1 Dominant and auxiliary feedback in transcriptional feedback modules**

---

In summery, GAT1 and DAL80 are found conserved in almost all the species of the Saccharomycetaceae family. The homologs of DAL80 and GAT1 in these species are also found to take part in nitrogen catabolism. Hence it is highly likely that the 'dominant' feedback mechanism of budding yeast NCR module is also conserved in the yeast species under the Saccharomycetaceae family.

### **5.1.7 Carbon catabolite repression (CCR) circuit**

Our argument of differential functioning of feedback loops in a module and conservation of the most vital one is more evident in several other modules, such as the carbon catabolite repression (CCR) circuit, which is the transcription factor sub-network controlling glucose uptake and utilisation. The carbon catabolite repression network, otherwise known as glucose repression network in budding yeast is a complex assembly of multiple signal transduction pathways and interweaved transcription regulatory networks. The whole glucose repression mechanism of yeast can be divided into three functional modules ([Kaniak et al. \(2004\)](#)). One of the modules operates through the Snf1 protein kinase and finally represses gene expressions at high level of glucose concentration. Another module senses the glucose signal via Snf3 and Rgt2 and induces the expression of glucose transporter genes. A third module uses Gpr1 and cAMP as messengers. Since the first two modules functions primarily by regulating gene expressions we focus our attention on these two modules. A representative diagram of these modules is shown in figure

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

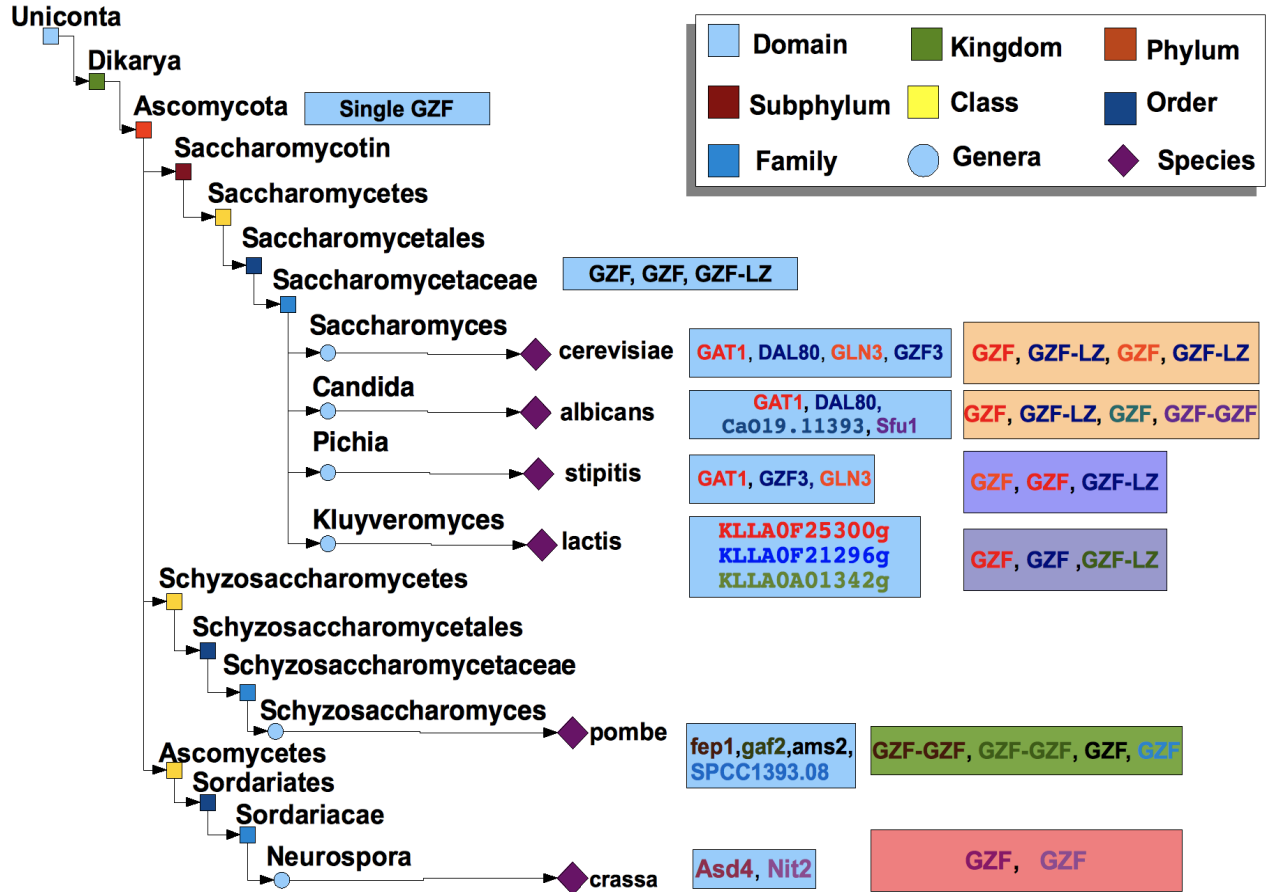


Figure 5.9: Evolution of GATA factor genes in fungi. Same colours indicate homologous genes. For example Gat1 in *S. cerevisiae*, *C. albicans*, and *P. stipitis*, and KLLA0F25300g in *K. lactis* are homologous and so on. The domain organisation is shown in a box right to the gene names and colour coded. For example, GAT1 and GZF are written in same colour. The species of Saccharomycetaceae family have at least one homolog of GZF(GATA binding Zn finger) and one of GZF-LZ(GATA binding Zn finger with an additional Leucine Zipper). However, species of other families contain double GZF-GZF encoding genes or many single GZF genes without any LZ. Hence, it is predicted that the ancestral yeast had only one GZF encoding genes and the other combinations appeared later in evolution. The data shown in this figure is collected from [Rossa and Peter \(2005\)](#) and [Low and Atchly \(2000\)](#).

### 5.10(a).

High level of glucose concentration is sensed by Snf3 and Rgt2 mediated pathways. In presence of glucose both Snf3 and Rgt2 binds with Mth1 and Std1 thus resulting in dilution of these two proteins in the cytoplasm. Mth1 and Std1 form a three protein complex with Rgt1 and the new complex represses the glucose transporter genes (HXTs). Once expressed the glucose transporters transport extracellular glucose into the cell and initiate glucose metabolism which in effect inhibits Snf1 triggering the repression of a wide range of genes ( Kaniak et al. (2004)). The genes which are regulated by these pathways regulate each other, thus weaving a complex regulatory pattern. Several feedback mechanism are found coupled and nested within each other in this transcription regulatory network. We extracted the transcriptional interplay of this module (see figure 5.10(b)) in order to study its kinetic behaviour.

As shown in figure 5.10(b) the CCR network in *S. cerevisiae* has nested feedback loops involving MIG1, MIG2, SNF3 and MTH1, a homolog of STD1 (Kaniak et al. (2004)). We develop mathematical model for the budding yeast CCR network to simulate its dynamics and understand its autonomous kinetic properties.

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

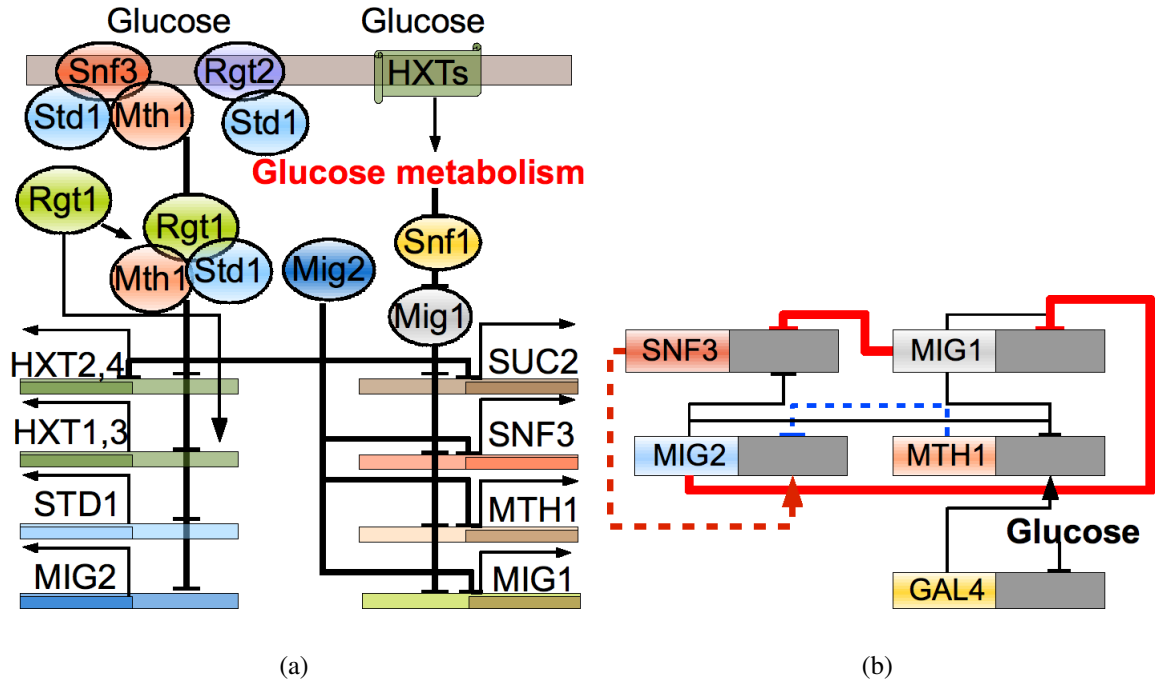


Figure 5.10: The glucose repression circuit of budding yeast (Kaniak et al. (2004)). The transcription regulatory feedback mechanisms of fig 5.10(a) are shown in figure 5.10(b). In figure 5.10(b), the arrowheads represent activation and the blunt arrows represent repression. The black solid lines represent direct transcription regulation and the blue dashed lines represent indirect transcription regulation via chains of protein interactions. In this case, both indirect transcription regulations are glucose dependent.

### 5.1.8 Dynamics of budding yeast CCR network

We developed a minimal mathematical model (equation 5.4) reflecting the transcriptional interplay of the genes which take part in the feedback module embedded in the glucose repression network. The model ignores many detailed mechanisms but captures the basic genetic regulations and is able to reproduce the qualitative behaviour of the corresponding genes.

$$\begin{aligned}
 \frac{dm_1}{dt} &= a_1 \frac{1}{1 + \left(\frac{m_1}{k_{11}}\right)^n + \left(\frac{m_2}{k_{12}}\right)^n} - \alpha_{m_1} m_1 \\
 \frac{dm_2}{dt} &= a_2 \frac{\left(\frac{s_3}{k_{21}}\right)^n}{1 + \left(\frac{m_3}{k_{22}}\right)^n + \left(\frac{s_3}{k_{21}}\right)^n} - \alpha_{m_2} m_2 \\
 \frac{dm_3}{dt} &= a_3 \frac{1}{1 + \left(\frac{m_1}{k_{31}}\right)^n + \left(\frac{m_2}{k_{32}}\right)^n + \left(\frac{g_1}{k_{33}}\right)^n} - \alpha_{m_3} m_3 \\
 \frac{ds_3}{dt} &= a_4 \frac{1}{1 + \left(\frac{m_1}{k_{41}}\right)^n + \left(\frac{m_2}{k_{42}}\right)^n} - \alpha_{s_3} s_3 \\
 \frac{dg_1}{dt} &= -\alpha_{g_1} g_1
 \end{aligned} \tag{5.4}$$

Here  $m_1$ ,  $m_2$ ,  $m_3$ ,  $s_3$  and  $g_1$  represent the concentrations of the products of MIG1, MIG2, MTH1, SNF3 and extracellular glucose. The repression of MTH1 by glucose is not direct as modelled in equation 5.4, instead, glucose represses GAL4 which in turn activates MTH1 resulting in a glucose mediated MTH1 repression. In reality, the concentration of extra cellular glucose depends on the HXTs which transport glucose into cell. Instead of modelling the whole mechanism in detail we have only simulated the changes in extracellular glucose concentration using simple first order decay law in order to simulate the effect of glucose



## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

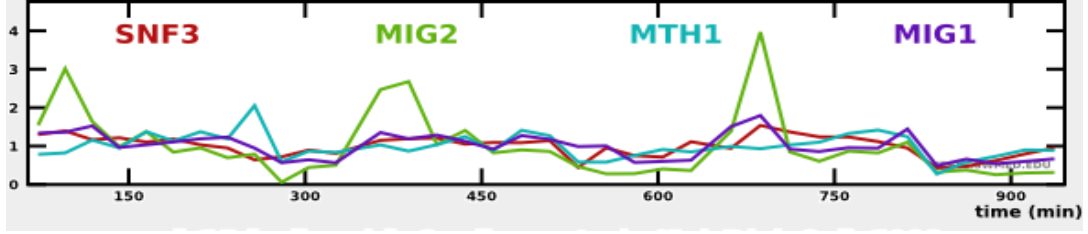
concentration on the expression of MTH1 gene. However, the kinetic behaviour of this network is fairly insensitive to the effect of glucose on MTH1 via GAL4. In equation. 5.4,  $a_i$  are the affinity factors of corresponding genes,  $\alpha_x$  are decay constants,  $k_{ij}$  are the Michaelis constants and  $n$  is hill coefficient which is set to 2 following common convention. The glucose repression pathway in budding yeast cross-talks with many pathways such as the cAMP pathway, the heat shock response pathway, the galactose induction pathway etc. (Kaniak et al. (2004)). Hence, to simulate the response of this pathway all the pathways which it crosstalks with need to be modelled in fair detail. Mathematical modelling of such a large system is difficult and needs sophisticated biological experiments to gather necessary data. We used a minimal modelling approach which does not take many details of the pathway into account. Hence, the parameter values of the pathway are chosen to qualitatively reproduce its behaviour as observed by Tu et al. (2005a) and are shown in table 5.3.

Params	Values	Params	Values	Params	Values
$a_1$	1.1296	$k_{22}$	0.0167	$\alpha_{m_3}$	0.303
$k_{11}$	0.059	$\alpha_{m_2}$	0.084	$a_4$	0.6086
$k_{12}$	0.045	$a_3$	0.5859	$k_{41}$	0.05675
$\alpha_{m_1}$	0.182	$k_{31}$	0.245	$k_{42}$	0.03436
$a_2$	0.195	$k_{32}$	0.07	$\alpha_{s_3}$	0.045
$k_{21}$	0.077	$k_{33}$	0.07	$\alpha_{g_1}$	0.0153

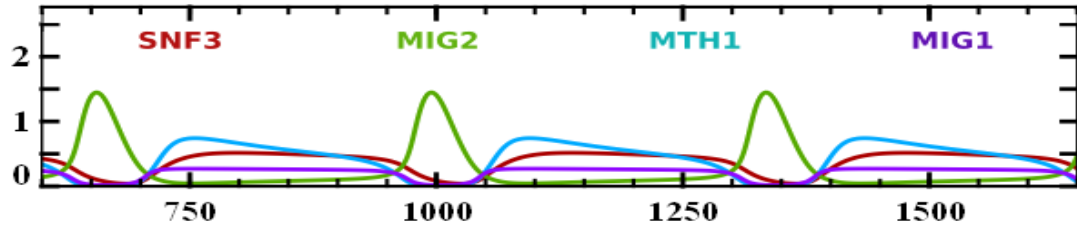
Table 5.3: Parameter values of equation 5.4 which represents the mathematical model of Carbon Catabolite Repression Circuit of wild type yeast cells.

The simulation of the model described in equation 5.4 along with the mRNA concentrations of the relevant genes as observed by Tu et al. (2005a) are shown in figure 5.11.

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules



(a)



(b)

Figure 5.11: Simulation of the mathematical model of carbon catabolite repression circuit as shown in equation 5.4. 5.11(a) shows the mRNA levels normalised to unit average as found by Tu et. al. [Tu et al. \(2005a\)](#). 5.11(a) is generated by SCEPTRANS [Kudlicki et al. \(2007\)](#). 5.11(b) shows the result of simulation of our model.

Bifurcation analysis of the wild type CCR model shown in equation 5.4 shows that it can produce sustained oscillation for a wide range of parameter values. The results of bifurcation analysis is shown in figure 5.12. It can be seen from figure 5.12 that the CCR network produces sustained oscillation between flipping the states of its genes from on and off. The switching occurs due to changes in its parameters which in many cases are glucose dependent. Hence, our analysis suggest that the CCR circuit produces sustained oscillations when the cell switches between different sources of sugars (glucose, sucrose, fructose, galactose etc.).

The period of the oscillation also varies between 60 and 350 mins

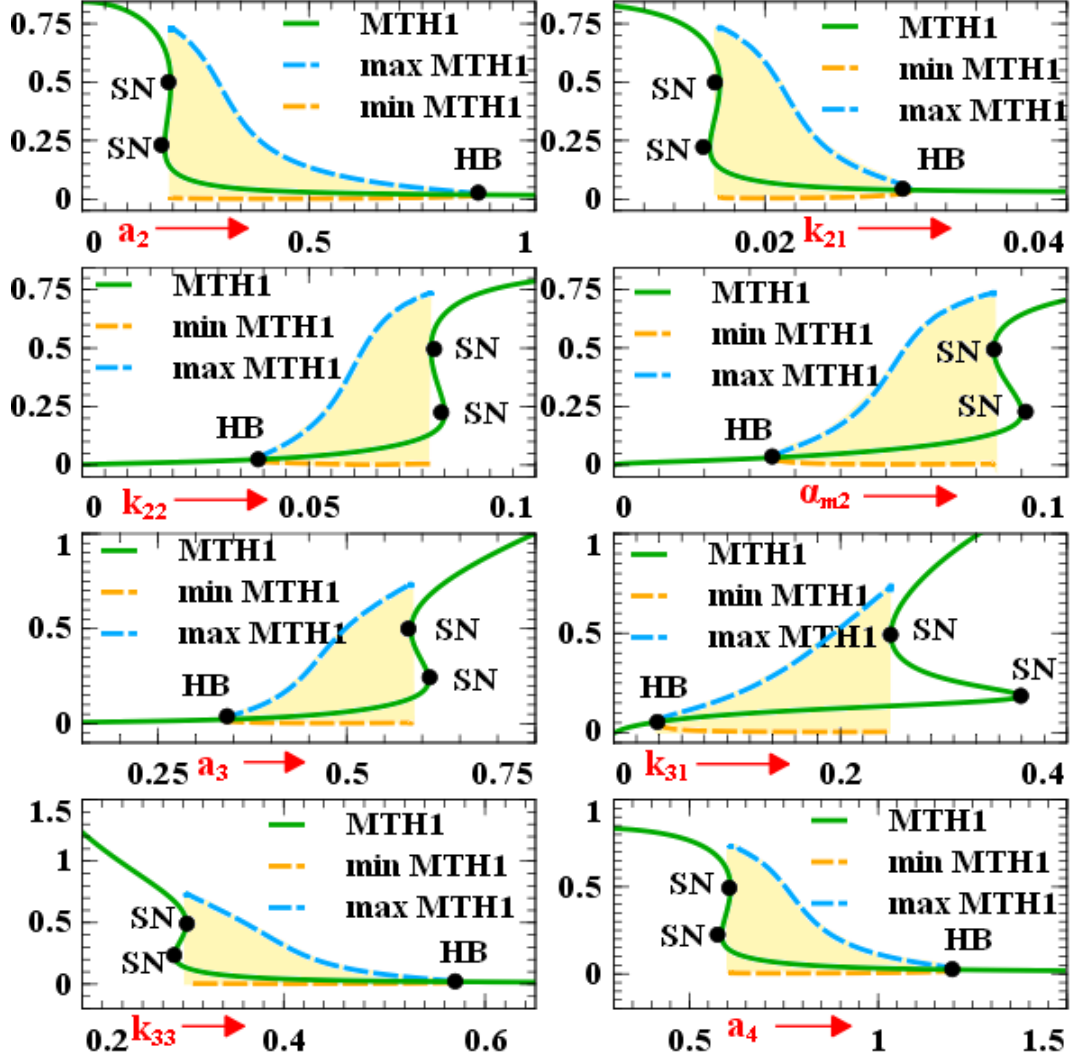


Figure 5.12: Bifurcation analysis of the dynamic behavior of the *S. cerevisiae* CCR circuit (only MTH1 is shown) due to perturbations in its parameters (see equation 5.4, table 5.3). Both saddle node(SN) and Hopf bifurcation(HB) are observed indicating that it can show bistable behaviour and oscillations. The genes in CCR network can exhibit sustained oscillations between 'on' and 'off' states due to changes in its parameters, some of which are glucose dependent (e.g.,  $k_{21}$ ,  $k_{22}$ ). The maximum and minimum concentrations during oscillation are shown by dashed lines. The oscillatory regions in the bifurcation diagrams are highlighted in yellow

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

---

depending on the parameter values which varies between different yeast strains and growth conditions. The period of oscillation is plotted against some of the parameters of the wild type budding yeast CCR model in figure 5.13. The timeframe is reminiscent of the ultradian clock that controls yeast metabolic cycles (Reinke and Gatfield (2006)). Although the exact cycle time is controversial and seems to depend on growth conditions and yeast strain, the existence of these metabolic cycles, where the cell switches between oxidative phosphorylation and anaerobic glycolysis, is well documented (Reinke and Gatfield (2006)). Tu et al. (2005a) suggested that such metabolic cycles helps to compartmentalise biological process temporally and provides better time management in cellular machinery.

### 5.1.9 MTH1 as an auxiliary gene

In the CCR circuit Rgt1 protein interacts with Std1 and Mth1 to form a three protein complex which represses the glucose transporter genes (HXTs). STD1 is homologous to MTH1 and have similar functions. According to Flick et al. (2003); Lakshmanan et al. (2003) a  $std1\Delta$  or a  $mth1\Delta$  mutant does not produce any lethal phenotype whereas a  $std1\Delta mth1\Delta$  double mutant completely abolishes the Rgt1 mediated repression of the HXT genes and renders the pathway useless. In other words, the carbon catabolite repression circuit of budding yeast needs only one of STD1 or MTH1 to operate satisfactorily. According to Jeffrey and Brown (2009) the CCR pathway of other yeast species such as

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

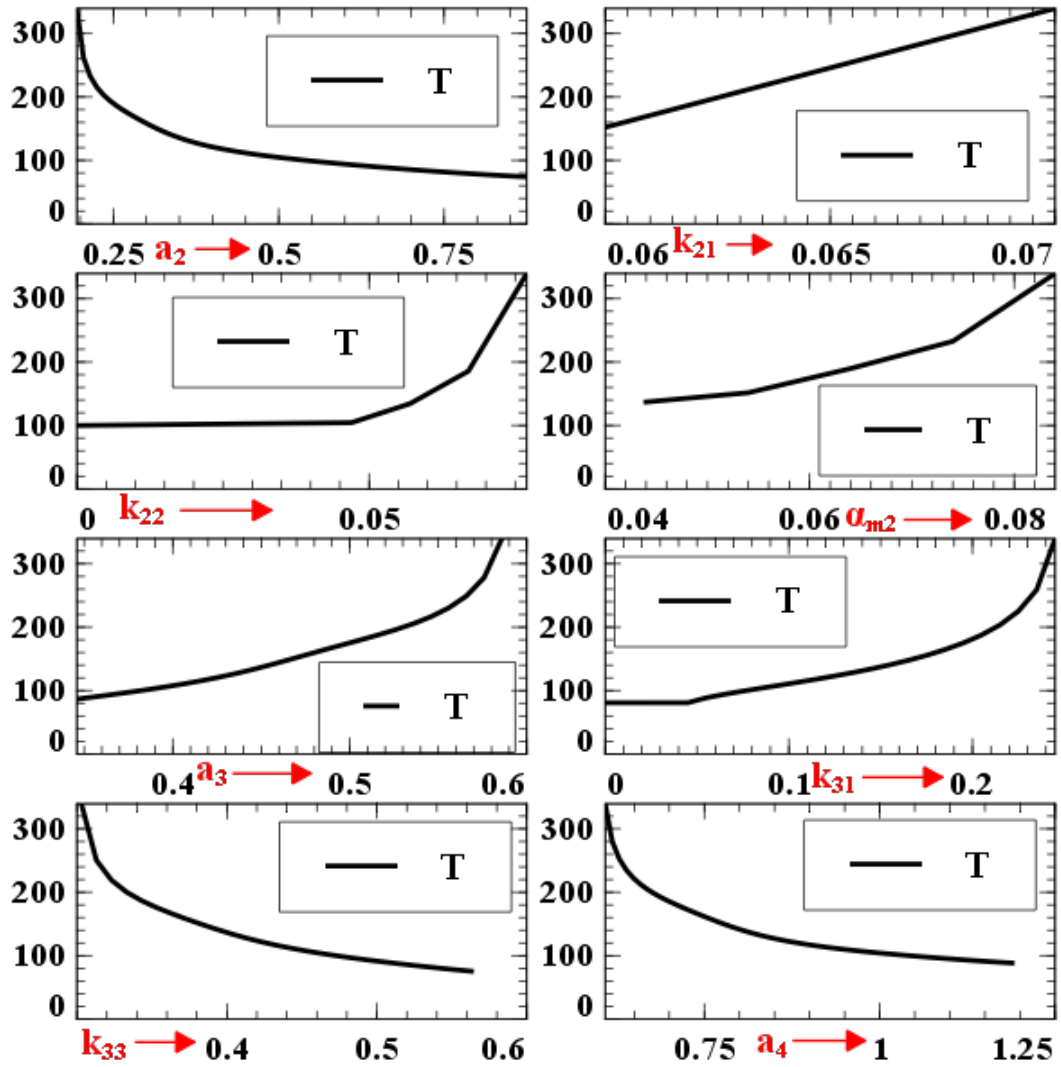


Figure 5.13: The variations in period of oscillation ( $T$ ) due to changes in parameter values. The period ( $T$ ) can vary from 60 mins to 350 mins depending on the parameter values.

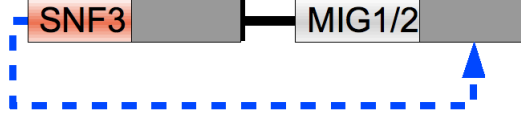


Figure 5.14: The transcriptional feedback mechanism of a  $mth1\Delta$  mutant. The  $mth1\Delta$  mutant has only one feedback loop made of SNF3 and MIG1/2

*C. albicans* has only one homolog of STD1 and lacks the extra MTH1. Interestingly, presence of only one STD1 homolog in the pathway considerably simplifies its structural complexity and reduces the feedback cluster into one transcriptional feedback mechanism. Hence, we investigate the roles of the MTH1 gene and the feedback loops appear due to its presence in the operation of the budding yeast CCR pathway by analysing the dynamic behaviour of a  $mth1\Delta$  mutant CCR circuit and comparing it with that of a wild type one.

### 5.1.10 Dynamics of $mth1\Delta$ mutant

The transcriptional feedback mechanism of a  $mth1\Delta$  mutant is shown in figure 5.1.10. The mathematical model of a  $mth1\Delta$  CCR circuit is shown in equation 5.5.

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

---

$$\begin{aligned}
\frac{dm_1}{dt} &= a_1 \frac{1}{1 + \left(\frac{m_1}{k_{11}}\right)^n + \left(\frac{m_2}{k_{12}}\right)^n} - \alpha_{m_1} m_1 \\
\frac{dm_2}{dt} &= a_2 \frac{\left(\frac{s_3}{k_{21}}\right)^n}{1 + \left(\frac{s_3}{k_{21}}\right)^n} - \alpha_{m_2} m_2 \\
\frac{ds_3}{dt} &= a_4 \frac{1}{1 + \left(\frac{m_1}{k_{41}}\right)^n + \left(\frac{m_2}{k_{42}}\right)^n} - \alpha_{s_3} s_3 \\
\frac{dg_1}{dt} &= -\alpha_{g_1} g_1
\end{aligned} \tag{5.5}$$

Bifurcation analysis of the  $mth1\Delta$  model reveals that the CCR pathway can not produce sustained oscillation without the MTH1 gene. MTH1 forms a three gene negative feedback loop in the wild type CCR pathway which is capable of producing sustained oscillation. This feedback is disrupted in a  $mth1\Delta$  mutant which results in inability of producing oscillatory response.

### 5.1.11 Dominant and auxiliary feedback in the CCR pathway of budding yeast

The above analysis suggests that the MTH1 gene in budding yeast CCR pathway helps to produce sustained oscillation when switching between carbon sources. Such oscillation is beneficial (as argued by [Tu et al. \(2005b\)](#)) to the metabolism of yeast but not essential since a  $mth1\Delta$  mutant produces no lethal phenotype. On the other hand,  $snf3\Delta$  mutant unable to grow in medium with low glucose concentration. Hence, we

## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

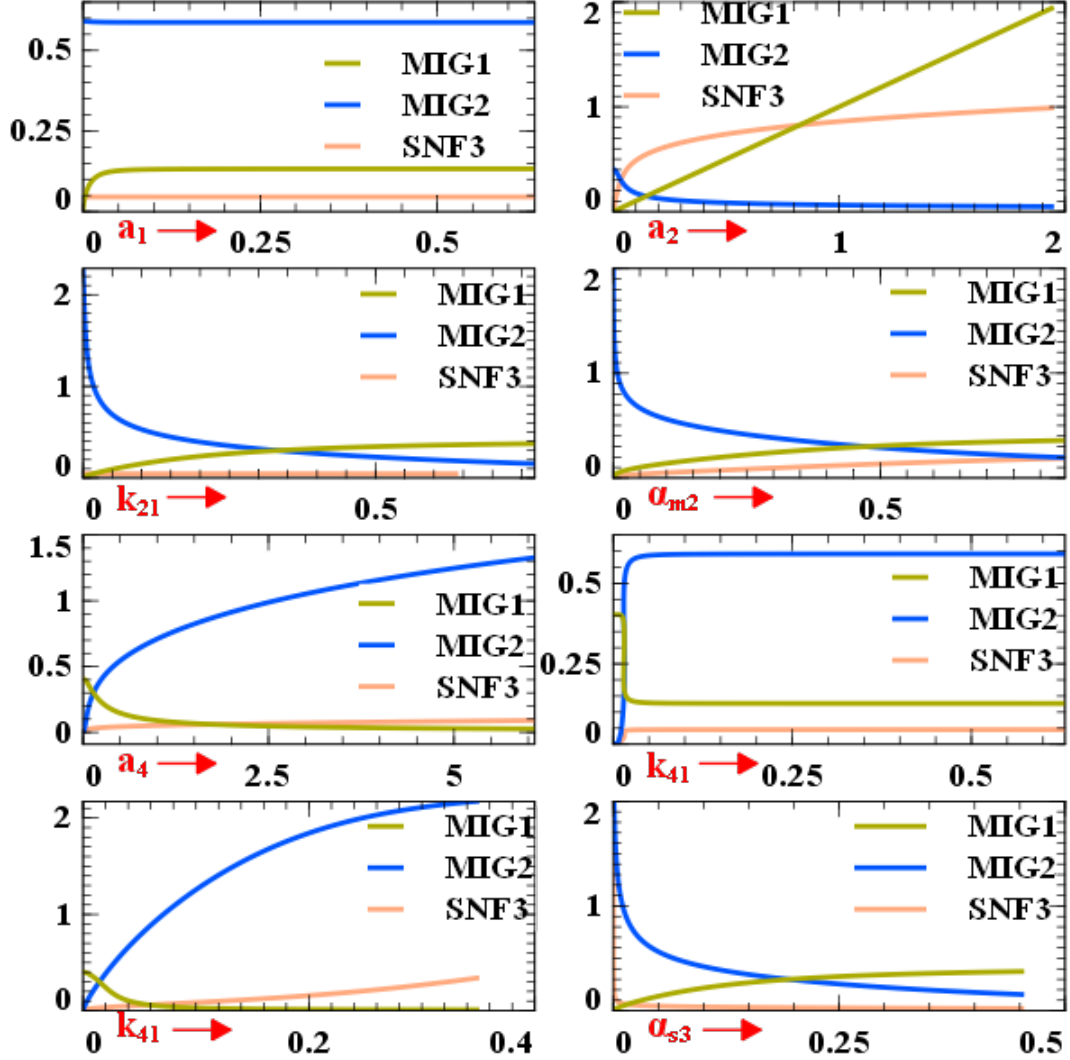


Figure 5.15: Bifurcation analysis of *mth1Δ* mutant CCR network of budding yeast. No bifurcation occurs for a wide range of parameter values in the neighbourhood of the operating point. Hence the *mth1Δ* mutant is unable to produce sustained oscillation.



## 5.1 Dominant and auxiliary feedback in transcriptional feedback modules

---

conclude that the transcriptional feedback between SNF3 and MIG1/2 acts as the main feedback in the CCR network whereas those mediated by MTH1 provides auxiliary functionality.

### 5.1.12 Evolution of carbon catabolite repression circuit in different yeast species

A complete picture of glucose repression mechanism is still not available for most of the yeast species other than budding yeast. However, a recent work by Jeffrey and Brown (2009) provides the most complete picture of the carbon catabolite repression of a common human pathogen *Candida albicans*. *C. albicans* is a yeast species which diverged from *S. cerevisiae* around 500 million years ago. A schematic diagram of *C. albicans* CCR network is shown in figure 5.16. It can be seen from figure 5.16 that the entire budding yeast CCR pathway is conserved in *C. albicans* apart from the interactions related to the MTH1 gene. Both *S. cerevisiae* and *C. albicans* belong to the Saccharomycetaceae family. Hence, we conclude that a basic mechanism which consists of a feedback loop between SNF3 and MIG1/2 might be conserved in the yeast species which belong to Saccharomycetaceae family. To understand the exact nature of the CCR evolution in yeast more experiments need to be done with other yeast species.

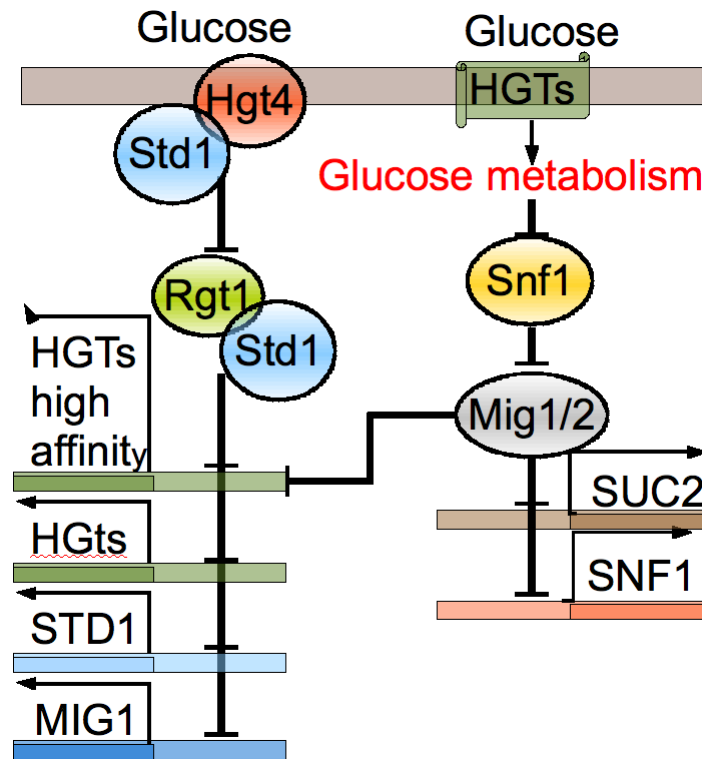


Figure 5.16: Glucose repression pathway of *Candida albicans* (Jeffrey and Brown (2009)). In this figure we have shown only part of the pathway which forms the transcriptional feedback mechanism. Some of the other interactions shown by Jeffrey and Brown (2009) are not shown in this figure for the convenience of visualisation.

## 5.2 A critical assessment of the analysis presented in this chapter

The analysis presented in this chapter is based on mathematical modelling of biochemical pathways using differential equations. This type of modelling approach comes with several limitations. Firstly, differential equation models are parameter rich and may exhibit similar transient properties for different sets of parameters Zenker et al. (2007). Similar transient properties do not guarantee similar steady state prop-

## 5.2 A critical assessment of the analysis presented in this chapter

---

erties. Hence, accurate estimation of parameters is necessary for correct determination of steady states properties. Parameter estimation of mathematical models of biochemical systems is a non trivial problem. Earlier methods of parameter estimation used different optimisation techniques to obtain a point estimate of the relevant parameter set [Schmidt and Jirstrand \(2006\)](#). More recent techniques [Girolami and Calderhead \(2011\)](#) estimates posterior probability over parameter values. Since bifurcation analysis needs a point estimate of parameter values we used the method implemented by [Schmidt and Jirstrand \(2006\)](#). We carried out mono-parameter bifurcation analysis in our study which computes the change of steady states due to perturbation of a particular parameter keeping the rest of the parameters fixed. This kind of analysis is called local bifurcation analysis since they depend on the point estimate of system parameters. A different set of point estimate which can also produce desired transient behaviour with reasonable accuracy may or may not lead to the same bifurcation properties. Since, the ODE models of biochemical pathways are usually highly non linear often their parameters have multimodal posterior distributions given a set of desirable transient responses. Multimodal distributions indicates the possibility of multiple set of parameters which can produce same transient response. Since our analysis is based on only the most likely point estimate and ignore other highly likely parameter values the bifurcation properties derived in this analysis are local. Hence the steady state behaviours shown in this analysis is one of many possibilities that can arise from the same pathways and should be treated as hypothesis rather than fact.

## 5.3 Concluding remarks

The two examples discussed in this chapter provide support to our hypothesis that in transcriptional modules consisting of multiple feedback mechanisms the biologically vital or the 'dominant' feedback loops are conserved and the 'auxiliary' feedback loops are not. The above conclusion is summarised in Fig. 5.17. We find that transcription regulatory feedback modules are usually small in *S. cerevisiae* typically spanning 2-5 genes. In protein interaction network however, feedback loops ranging from small auto-regulatory units to a few hundred proteins are coupled together (see table 3.2) and form complex clusters. This raises many interesting questions, such as: how did evolution shape such organisation of feedback mechanisms in yeast biochemical networks? Which of these feedbacks have dominant roles in the operation of the module and why? We approach these questions by analysing the dynamic behaviour of some well studied feedback mechanisms which appear in the stress response and mating pathways of *S. cerevisiae*. Here, our goal is to determine the role of different types of feedback mechanisms involved in these modules and classify them into 'dominant' and 'auxiliary' loops. Based on these data we can then examine the conservation of both types of feedback loops in other yeast species.

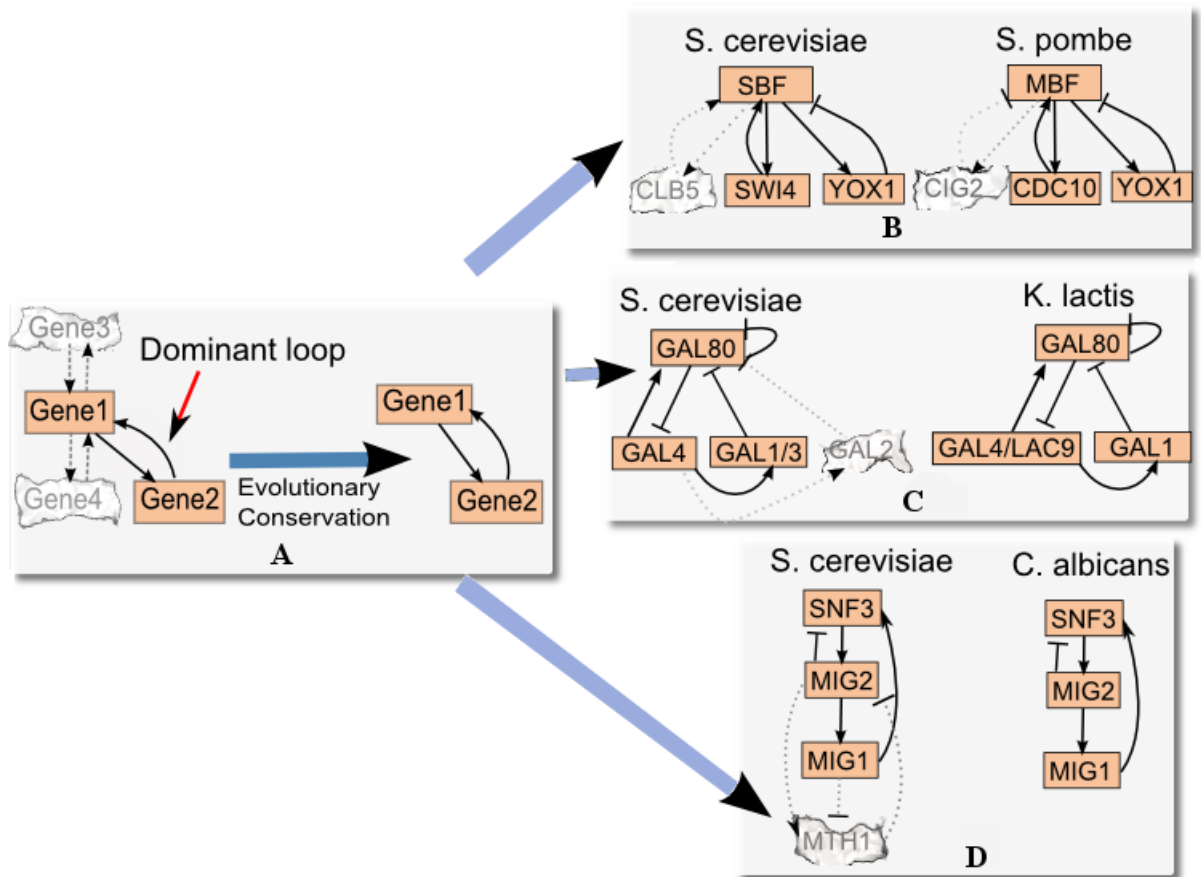


Figure 5.17: Biologically vital or functionally 'dominant' feedback loops in a transcriptional feedback module are most likely to be conserved. (a) A conceptual diagram of evolutionary conservation in a four gene feedback module. The functionally dominant loop is drawn as solid line and remains conserved during evolution. Auxiliary loops are drawn as thin broken lines and are less conserved. (b) shows the preferential conservation of the possibly 'dominant' loops in a cell cycle related transcriptional feedback modules of *S. cerevisiae* and *S. pombe*. (c) shows another example of evolutionary conservation of the GAL80-GAL4-GAL1/3 feedback module in *S. cerevisiae* and *K. lactis*. The GAL2 mediated feedback in the same module of *S. cerevisiae* is not conserved in *K. lactis*. (d) shows the conservation of functionally 'dominant' feedback loops of the carbon catabolite repression module in *S. cerevisiae* and *C. albicans*. The MTH1-MIG1/2 mediated auxiliary feedback loop is not conserved in *C. albicans*.

## Chapter 6

# Study of feedback modules in protein interaction networks of different yeast species

According to [Yu and Gerstein \(2006\)](#) yeast transcription regulatory networks are hierarchical with four levels of hierarchy, i.e. long chains of transcription regulations are rare. This is not the case in protein interaction networks. In yeast long chains of protein interactions are common in metabolic networks ([Herrgard et al. \(2008\)](#)), signalling cascades ([Posas et al. \(1998\)](#)), cell cycle network etc. Hence some of the feedback mechanisms found in protein interaction networks are longer than those normally found in transcription regulation network of yeast. Additionally, they often enclose small feedback loops in their paths, e.g. pheromone response pathway ([Paliwal et al. \(2007\)](#)), osmoregulation

---

pathway (Klipp et al. (2005)) etc. Such symbiotic organisation of long and small feedback loops in protein interaction networks raises many interesting questions, e.g., which feedback loops are biologically vital in these clusters, the longer ones or the smaller ones nested within them? Attempt to answer this question may lead to two different school of thoughts.

- Firstly, it is possible that the small feedback mechanisms nested within the longer ones are biologically vital and the long ones enclosing them provide protection from stochastic noise.
- On the other hand, it's also possible that the long feedback mechanisms are biologically vital where the small ones provide extra modulation to their operation.

Answering this question may lead us to find the evolutionary trend of conservation in feedback modules of protein interaction networks. For example, if the long range feedback loops are biologically vital then it makes more sense to assume that these feedback loops are evolutionarily conserved (and vice versa) since disrupting the long cascades might produce deleterious phenotypes. To select between the two hypothesis discussed above we analyse the dynamics and evolution of few well studied pathways which have nested feedback architecture, such as, the pheromone response pathway and the osmoregulation pathway of yeast. To remain consistent with the previous chapter we shall call the biologically most vital feedback loops as 'dominant' loops and the other feedback loops as 'auxiliary' loops. Our goal in this chapter is to deter-

mine the 'dominant' and 'auxiliary' loops of pheromone response and osmoregulation pathways and look for their evolutionary conservations in different yeast species.

### 6.1 Pheromone response pathway

Yeast pheromone response pathway is a combination of long chains of MAPK cascades and transcription regulations of several genes. It directly participates in yeast chemotropism in pheromone gradients (Paliwal et al. (2007)). This pathway senses pheromone gradient and helps mating of two haploid cells of opposite mating types which results in the formation of a diploid zygote. Several studies have investigated this pathway in order to understand its operating procedures and provided many useful insights, e.g. Kofahl and Klipp (2004) explained its graded dose response and Paliwal et al. (2007) explained the switch-like behaviour of this pathway. In this section we shall re-visit some of these studies in order to understand the relation between its structural organisation and dynamical behaviour which will eventually lead us to understand the plausible evolutionary trends in preserving this pathway in other species.

In yeast pheromone response pathway, a long chain MAPK cascade made of Ste2/3, Ste20, Ste11, Ste7, Fus3, Ste12, Bar1 and Cdc25 is activated when the pheromone sensors Ste2/3 sense high pheromone



concentration in the growth medium (Kofahl and Klipp (2004)). The pheromone signal is carried and amplified by the MAPK cascade which eventually phosphorylates Ste12. Phosphorylated Ste12 enters nucleus and activates its own transcription and that of FUS1, FUS3 and BAR1 (Paliwal et al. (2007)). Activated BAR1 then shuttles out of nucleus and eventually degrades pheromone (Kofahl and Klipp (2004)). Thus activation of the MAPK cascade by pheromone concentration and the degradation of pheromone molecules by the pheromone response pathway results in a long feedback mechanism which tracks the pheromone concentration in growth medium. This long feedback mechanism encloses a number of small feedbacks such as the auto-regulation of Ste12 and the positive feedback loops between Ste12, Fus3 and Kss1 (Paliwal et al. (2007)). A schematic diagram of the *S. cerevisiae* pheromone response pathway is shown in figure 6.1.

In this section we shall revisit the dynamic properties of pheromone response pathway. But unlike the earlier studies we shall rather focus on the contributions of different kinds of feedback loops in the dynamics of pheromone response pathway in order to determine its 'dominant' and 'auxiliary' loops. Finally we shall look for evolutionary conservation of both 'dominant' and 'auxiliary' loops of pheromone response pathway of different yeast species.

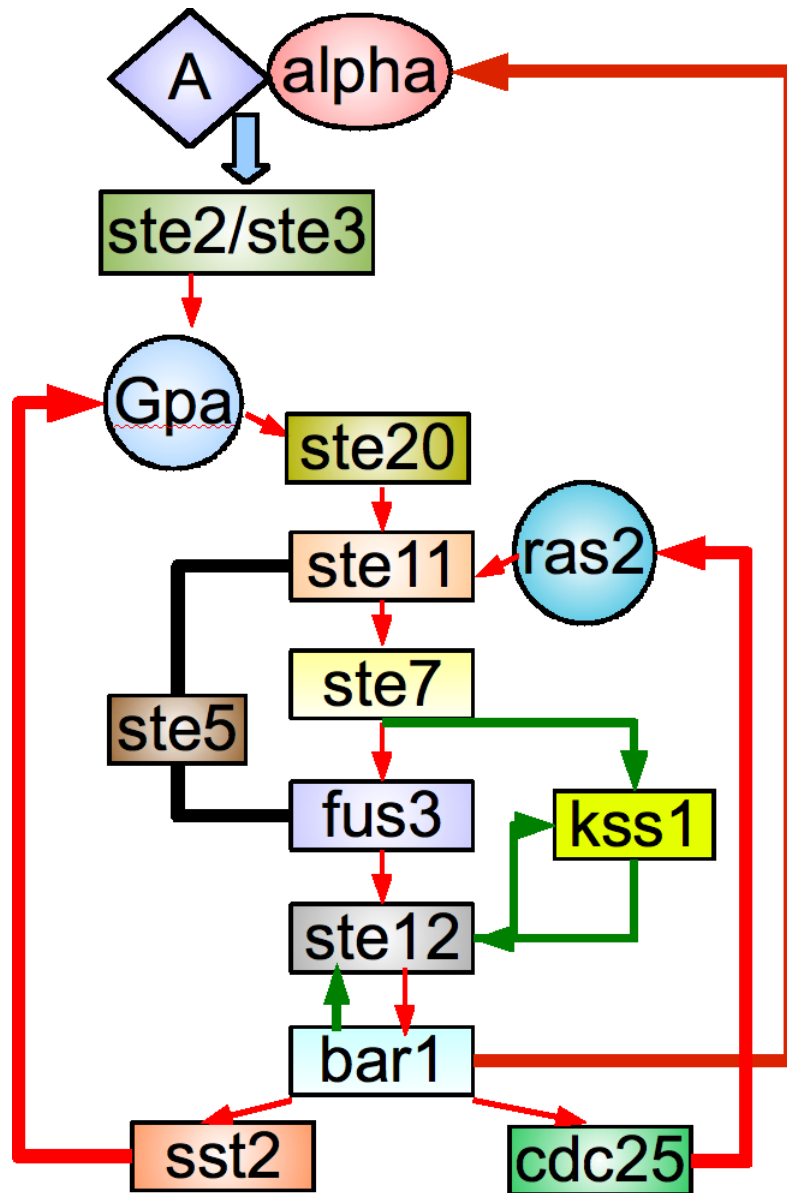


Figure 6.1: Schematic diagram of pheromone response pathway of budding yeast. This pathway has a long MAPK cascade which activates Bar1 and degrades pheromone resulting in a feedback regulation in pheromone sensing. The long feedback encloses many small transcriptional and proteomic feedback loops. Here we have shown only the feedback loops caused by protein interaction. Such as the positive feedback loop which arises due to the activation of Fus3 by Ste12 and vice versa. The self regulatory transcriptional feedback of Ste12 is not shown in this figure.

### 6.1.1 Dynamics of yeast pheromone response pathway

Several mathematical models of the yeast pheromone response pathway have been proposed so far. Among them, [Kofahl and Klipp \(2004\)](#)'s model presents one of the most detailed mathematical description of the signalling cascade that this pathway possess but it does not describe the transcriptional interplay between STE12, FUS3 and FUS1. Hence, their model accurately describes the graded response to pheromone gradient but is unable to explain the switch like behaviour observed by [Paliwal et al. \(2007\)](#). On the other hand, [Paliwal et al. \(2007\)](#)'s model is rather minimalistic and mainly focuses on the transcription regulations of the pheromone sensing mechanism. Additionally [Paliwal et al. \(2007\)](#)'s model does not take the Bar1 mediated pheromone degradation into account, thus eliminates the feedback mechanism which is responsible for the pheromone sensing in the first place. Introducing the Bar1 mediated interaction to [Paliwal et al. \(2007\)](#)'s model presents a rather complete picture but alters its entire state space. In [Paliwal et al. \(2007\)](#)'s model pheromone concentration is regarded as constant. However, introducing a biochemical interaction which continuously degrades pheromone is possible only if pheromone concentration is considered as a separate state. In this case the trajectory of the new state space might be entirely different compared to the original trajectory observed by [Paliwal et al. \(2007\)](#) due to introduction of a new state. We develop a mathematical model for the pheromone response pathway of wild type budding yeast by adding the Bar1 mediated feedback mechanism to the [Paliwal et al. \(2007\)](#)'s model to understand the dynamics of the complete system. The ordinary differential equation model we developed is shown in equation

6.1 and the parameters are shown in table 6.1. The first eight equations in 6.1 are taken from Paliwal et al. (2007) and the last three equations are taken from Kofahl and Klipp (2004) which includes the Bar1 mediated main feedback mechanism in the model.

$$\begin{aligned}
 \frac{dF}{dt} &= -k_{f1}F \frac{\alpha^{m1}}{k_{m1}^{m1} + \alpha^{m1}} + k_{r1}F_p + c_1\gamma + c_2\gamma \frac{S_p^{n1}}{c_3^{n1} + S_p^{n1}} - k_{d1} * F \\
 \frac{dF_p}{dt} &= -k_{f1}F \frac{\alpha^{m1}}{k_{m1}^{m1} + \alpha^{m1}} - k_{r1}F_p - k_{d2}F_p \\
 \frac{dK}{dt} &= -k_{f2}K \frac{\alpha^{m2}}{k_{m2}^{m2} + \alpha^{m2}} + k_{r2}K_p - k_{f4}K.S + c_4\delta + c_5\delta \frac{S_p^{n2}}{c_6^{n2} + S_p^{n2}} \\
 &\quad - k_{d3}K \\
 \frac{dK_p}{dt} &= k_{f2}K \frac{\alpha^{m2}}{k_{m2}^{m2} + \alpha^{m2}} - k_{r2}K_p - k_{d4}K_p \\
 \frac{dS}{dt} &= -k_{f3}SF_p - k_{f3_k}SK_p + k_{r3}S_p - k_{f4}KS + K_{r4}S_b + c + c_7 \\
 &\quad + c_8 \frac{S_p^{n3}}{c_9^{n3} + S_p^{n3}} - k_{d5}S \\
 \frac{dS_p}{dt} &= k_{f3}SF_p + k_{f3_k}SK_p - k_{r3}S_p + k_{d6}S_p - k_{spb1}B_1S_p \\
 \frac{dS_b}{dt} &= k_{f4}KS - k_{r4}S_b - K_{d7}S_b \\
 \frac{dR}{dt} &= c_{10} + c_{11} \frac{S_p^{n4}}{c_{12}^{n4} + S_p^{n4}} - k_{d8}R \\
 \frac{dB_1}{dt} &= -k_{spb1}B_1S_p + k_{fpb}B_{1p} - k_{d9}B_1 \\
 \frac{dB_{1p}}{dt} &= k_{spb1}B_1S_p - k_{fb}B_{1p} - k_{ab1p}\alpha B_{1p} - k_{d10} * B_{1p} \\
 \frac{d\alpha}{dt} &= B_\alpha - k_{ab1p}\alpha B_{1p} - k_{d11}\alpha
 \end{aligned} \tag{6.1}$$

## 6.1 Pheromone response pathway

Here  $F$  =inactive Fus3,  $F_p$ =active Fus3,  $K$  =inactive Kss1,  $K_p$  =active Kss1,  $S$  =inactive Ste12,  $S_p$  =phosphorylated Ste12,  $S_b$  = Ste12 repressed by Fus3, Dig1 and Dig2,  $B_1$ =inactive Bar1,  $B_{1p}$  =phosphorylated Bar1,  $\alpha$  = pheromone. The description and the values of the kinetic parameters are similar to those in [Kofahl and Klipp \(2004\)](#); [Paliwal et al. \(2007\)](#).

Params	Values	Params	Values	Params	Values	Params	Values	Params	Values
$k_1$	4	$c_{10}$	0.25	$k_{r3}$	30	$k_{d8}$	0.04	$m2$	2
$c_1$	4	$c_{11}$	3	$k_{f4}$	2.5	$k_{d9}$	0	$\gamma$	1
$c_2$	4	$c_{12}$	150	$k_{r4}$	5	$k_{d10}$	0	$\delta$	1
$c_3$	50	$c_{13}$	4	$k_{d1}$	0.04	$k_{d11}$	0.04	$k_{b1p}$	0.01
$c_4$	4	$k_{f1}$	100	$k_{d2}$	0.02	$n_1$	2	$k_{spb1}$	5
$c_5$	0	$k_{m1}$	1	$k_{d3}$	0.04	$n_2$	2	$B_\alpha$	0.01
$c_6$	50	$k_{r1}$	10	$k_{d4}$	0.02	$n_3$	3	$k_{ab1p}$	0.03
$c_7$	1	$k_{f2}$	20	$k_{d5}$	0.04	$n_4$	2	$k_\alpha$	0.01
$c_8$	4	$k_{f3}$	10	$k_{d6}$	0.02	$n_5$	2	$k_{fb}$	0.001
$c_9$	60	$k_{f3k}$	5	$k_{d7}$	0.04	$m1$	2	-	-

Table 6.1: Parameter values for equation 6.1 which represents the mathematical model of pheromone response pathway.

[Paliwal et al. \(2007\)](#) showed that their model exhibits bistability due to changes in pheromone concentrations and provided experimental evidences to confirm their claim. In our model pheromone concentration is a state itself, hence we analyse the trajectory of the system due to fluctuations in the basal rate of pheromone production. We considered such rate to be constant in our model. The basal rate of pheromone production represents the rate by which the yeast cells of opposite mating type produces pheromone. Like [Paliwal et al. \(2007\)](#) we also find that the pheromone model exhibits bistability due to fluctuations in the basal pheromone

production rate. Bistability results in hysteresis due to changes in the parameters of the pathway (see figure 6.2). Hysteresis is beneficial for the pheromone response pathway since it incorporates history dependent activation in its dynamic behaviour. When a yeast cell encounters high pheromone concentration its pheromone response pathway gets activated which degrades pheromone itself in the growth medium via Bar1. Fast pheromone degradation may dilute the pheromone concentration in the growth medium beyond a threshold level which might result in shut down of the pathway before successful cell cycle arrest. Due to history dependent activation the pathway remains activated long enough despite fast pheromone degradation to ensure successful cell cycle arrest and smooing. Hence the bistable behaviour of the system reduces the cost of expensive pheromone dependent mating mechanism. However, we also find that both Paliwal et al. (2007)'s model and our model exhibit bistability due to fluctuations in many other reaction rates that governs the system. A bifurcation diagram is shown in figure 6.2.

### 6.1.2 Dynamics of a $kss1\Delta$ mutant

Finally, to understand the role of KSS1 in the pheromone response pathway we analyse the dynamic behaviour of a  $kss1\Delta$  mutant and compare the analysis with that of the wild type one. Following Paliwal et al. (2007), we simulate a  $kss1\Delta$  mutant by making  $\delta = 0$ . Our analysis suggests that in  $kss1\Delta$  mutant hysteresis is either absent or too narrow to be beneficial for the operation of the system (see figure

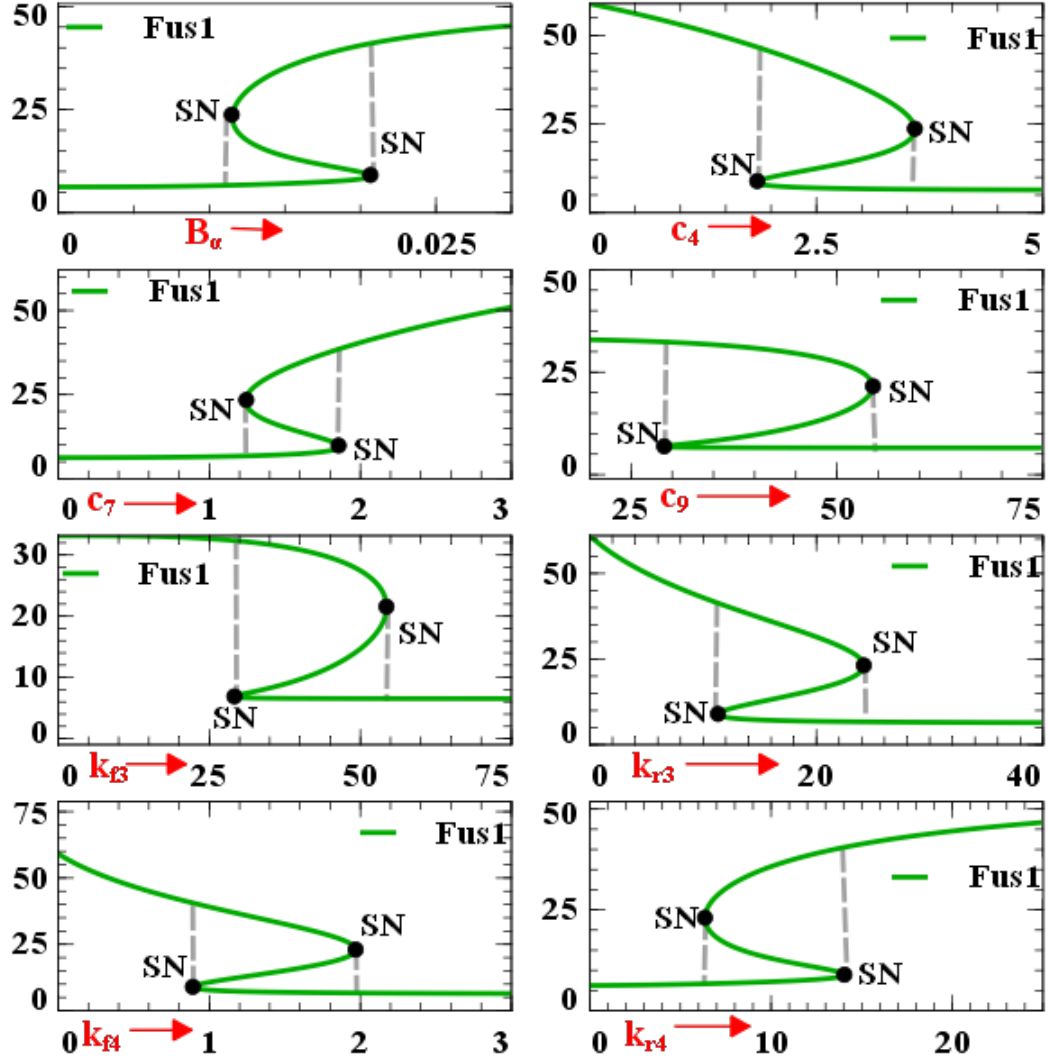


Figure 6.2: Bifurcation analysis of the mathematical model (equation 6.1, table 6.1) of the budding yeast pheromone response pathway. The model exhibits hysteresis and bistability (saddle node bifurcations, denoted by SN in the figure) due to fluctuations in many of its parameters including the basal rate of pheromone production  $B_\alpha$ . The results are similar to Paliwal et al. (2007).

6.3). Hence, the role of KSS1 in the pheromone response pathway is mainly to widen the distance between its stable states. The hysteresis produced by KSS1 is beneficial to the organism since it reduces the cost of pheromone dependent mating by making smooing possible more often than otherwise during pheromone tracking (Paliwal et al. (2007)). But such a system is not essential since a yeast cell survives without the KSS1 gene (Paliwal et al. (2007)).

### 6.1.3 Dominant and auxiliary feedback loops in pheromone response pathway

We suggest that the feedback loop made of the long chain of MAPK cascades and the transcription regulation of the FUS1,3 and STE12 'dominant' loop due to essential role in pheromone response whereas the small feedback mechanism made of Kss1 is a beneficial upgrade to the vital mechanism and can be termed as auxiliary loop. Hence, in this case, it is highly likely that the biologically most vital feedback mechanism is conserved among different yeast species since disruption in this mechanism may proved to be deleterious. On the other hand, the small feedback mechanism which occur due to the presence of Kss1 may not be conserved in other yeast species due to its rather non-essential role in pheromone sensing and tracking. To test this hypothesis we reconstructed the pheromone response pathway of fission yeast from published data and compared with that of budding yeast. The comparison of budding and fission yeast pheromone response pathways are given in



## 6.1 Pheromone response pathway

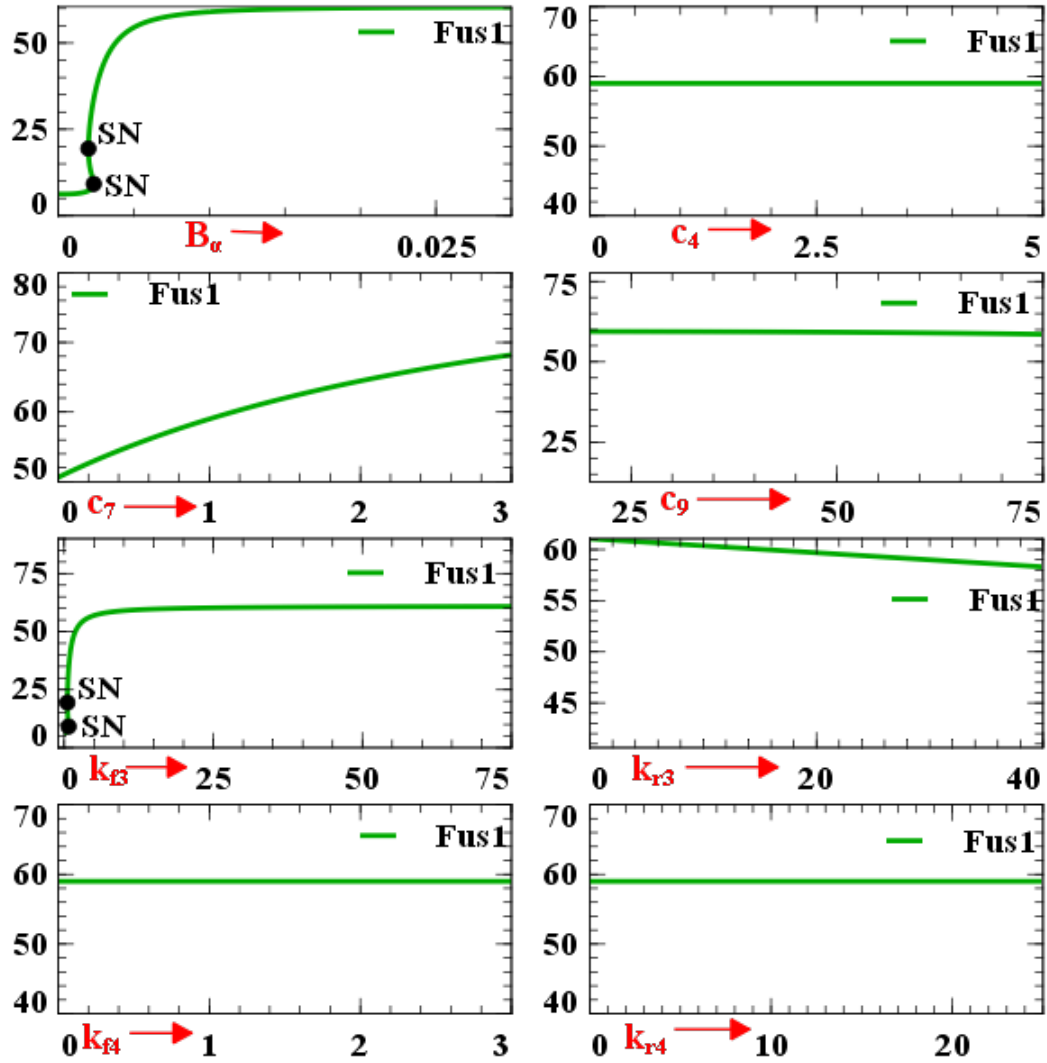


Figure 6.3: Bifurcation analysis of a  $kss1\Delta$  mutant. In a  $kss1\Delta$  mutant the hysteresis is either absent or too narrow to make a difference.

the next section.

### 6.1.4 Conservation of pheromone response pathway in other yeast species

We have reconstructed the pheromone response pathway of fission yeast to compare it with that of budding yeast. We used the interactions and homology information provided by [Herskowitz \(1995\)](#); [Hughes et al. \(1994\)](#); [Imai and Yamamoto \(1992\)](#); [Marcus et al. \(1995\)](#); [Vohra et al. \(2003\)](#) to reconstruct the fission yeast counterpart of yeast pheromone response pathway. For convenience of comparison we have shown the budding and fission yeast pheromone response pathways side by side in figure 6.4. It can be seen from figure 6.4 that the entire pathway is conserved in fission yeast apart from the small 'auxiliary' feedback loop resulted due to the presence of KSS1. Since KSS1 is homologous to FUS3 we presume that KSS1 appeared later in yeast family after the divergence of fission and budding yeast probably due to whole genome duplication events. There is only one homolog of FUS3 in *S. pombe* and hence the nested feedback loops do not exist. From the above analysis we find that in pheromone pathway the longest feedback loop is the 'dominant' loop and is conserved in even the most distant of its relatives. However, the small KSS1 mediated feedback is beneficial but not essential to the operation of the module and are rather species specific.

The above analysis points toward some interesting possibilities. For example, it is possible that in large feedback modules of yeast biochemical networks the long range feedback mechanisms might be functionally

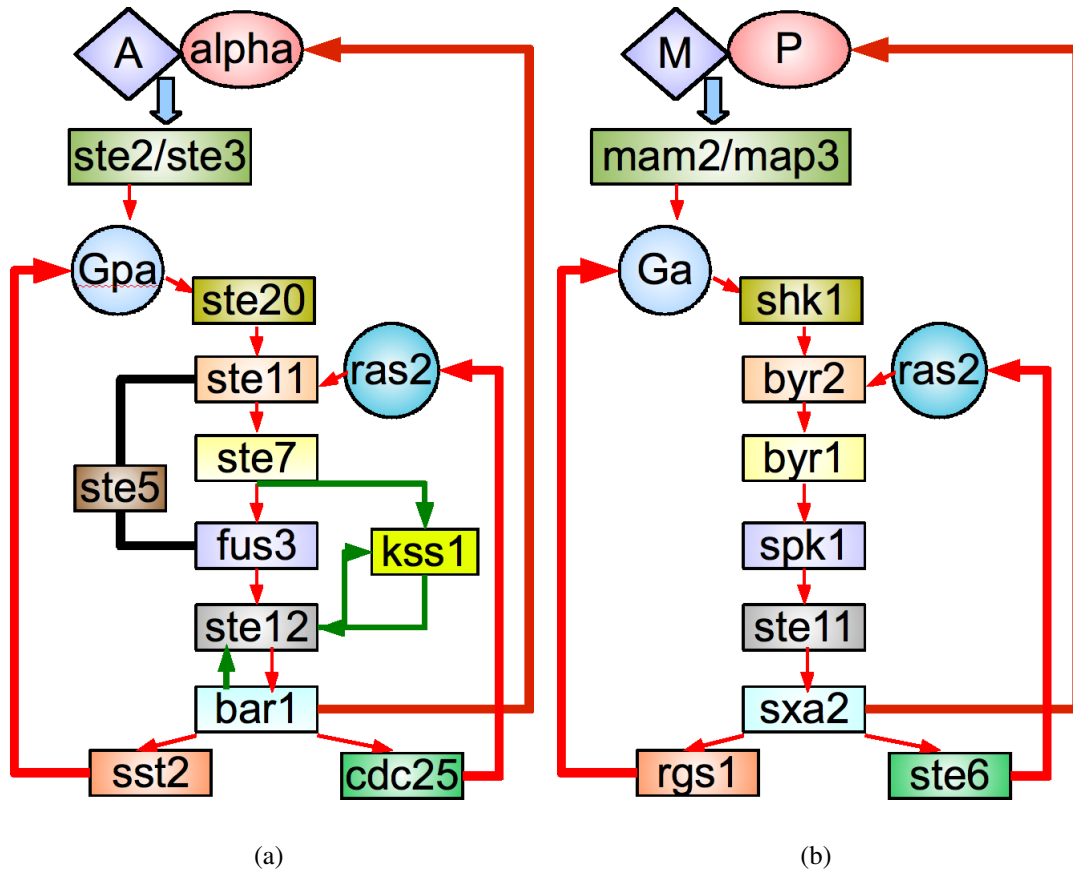


Figure 6.4: Pheromone response pathways in budding (a) and fission yeast (b). Boxes of same colours represent homologous genes. For example, mam2/map3, ste2/ste3;  $G_{\alpha}$ (budding yeast),  $G_{\alpha}$ (fission yeast); byr2, ste11; byr1, ste7; spk1, fus3/kss1 are homologous Herskowitz (1995). Similarly, ste12(budding yeast), ste11(fission yeast); rgs1, sst2 Hughes et al. (1994); Vohra et al. (2003); and sxa2, bar1; Imai and Yamamoto (1992) shk1, ste20 Marcus et al. (1995) are homologous.

'dominant', whereas the smaller nested ones might be 'auxiliary'. In that case, the longer feedback loops should be conserved in other yeast species, whereas the smaller ones may not be conserved. One plausible argument to support this hypothesis is as follows. Long range feedback loops are found mainly in signalling cascades which sense and respond to environmental stimuli. Therefore, these mechanisms need to function quickly and decisively. Previous studies [Nakabayashi and Sasaki \(2005\)](#) showed that the design of signalling cascades is usually optimal to achieve certain end product concentration at minimal possible response time. The optimal functionality of these cascades might be disrupted by any perturbation and need to be protected against stochastic noises. The feedbacks enclosing these cascades can provide such protection. Hence, these feedback loops remain conserved to protect their optimal functionality.

We decided to look for similar instances in other pathways of *S. cerevisiae*, such as the osmo-regulation pathway before coming to a conclusion.

## 6.2 Osmoregulation pathway

Osmoregulation pathway is another example where long and small feedback loops are found coupled with each other and assist in each others operation. When a yeast cell is subjected to excess osmotic pressure its

osmoregulation pathway produces glycerol to increase the cell turgor as a counter measure (Klipp et al. (2005)). In budding yeast, it uses two parallel pathways made of long chains of signalling cascades that converge at Pbs2 (Klipp et al. (2005)). Pbs2 further activates Hog1 which eventually produces glycerol and thereby completes a long feedback by buffering excess osmotic pressure. Hog1, however, inhibits itself via a small negative feedback loop mediated by the activation of Ptp2,3 which inhibit Hog1. See figure 6.5 for a schematic diagram of the Sln1 branch of the budding yeast osmoregulation pathway. In this section we analyse the dynamic behaviour of the budding yeast osmoregulation pathway to understand the role of ptp2,3 in its functioning.

### 6.3 Dynamics of wild type budding yeast osmoregulation pathway

Several mathematical model have been developed so far to simulate and predict the behaviour of the wild type osmoregulation pathway of *S. cerevisiae*. Some of these models are very detailed and accurate in reproducing the behaviour of this pathway and some of them are rather minimalistic but still captures the basic dynamics of the osmoregulation system. For example, the mathematical model developed by Klipp et al. (2005) captures most of the molecular details of the osmoregulation pathway in *S. cerevisiae*. On the other hand the models such as that of Gennemark et al. (2006) are minimalistic and does not captures all details

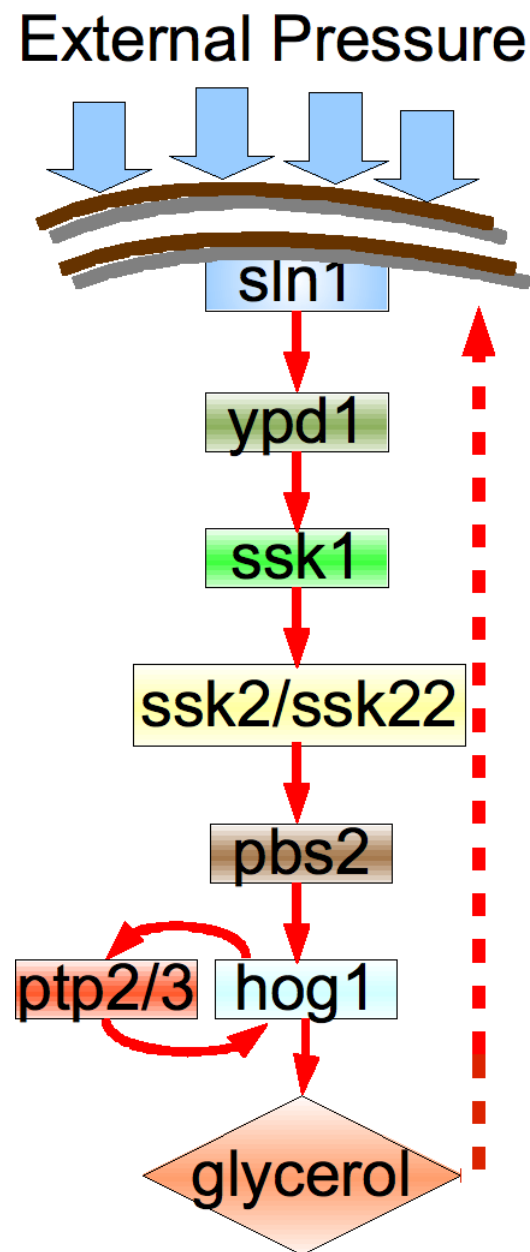


Figure 6.5: The osmoregulation pathway of budding yeast.

of the osmotic shock response mechanism. There are both advantages and disadvantages of both the modelling approach. Though detailed modelling are more accurate, the state space and parameter space of these models are extremely high dimensional which makes any computational analysis difficult. For example, high dimensional parameter space usually increases the computational cost of statistical inference of parameter values from time series data and high dimensional state spaces may destabilise numerical continuation algorithms in phase space. On the other hand, minimal mathematical models do not have such problems but they often miss important molecular details. Hence we developed a mathematical model which is simple enough for our computational purpose but complex enough to describe all necessary details.

Our model takes into account both the Sho1 and Sln1 mediated paths and includes a small feedback loop mediated by ptp2 (protein tyrosine phosphatase found in nucleus). It replicates the experimental conditions of [Krantz et al. \(2004a\)](#). In [Krantz et al. \(2004a\)](#)'s experiment the osmotic pressure of the growth medium of wild type yeast cells was increased(hyper-osmotic shock) by adding 0.5M NaCl which exerts an external osmotic pressure  $\Pi_e$  of approx 2.4 MPa. In a typical growth medium (osmotic pressure 1.4-1.5 MPa [Gervais et al. \(1996\)](#)) an average yeast cell has a turgor pressure  $\Delta P$  of approximately 0.18MPa ([Gervais et al. \(1996\)](#)). Hence the internal osmotic pressure  $\Pi_i$  of an average yeast cell is around  $1.4 + 0.18 \sim 1.6$  MPa. The difference between internal and external pressure ( $\Pi_e - \Pi_i$ ) is around  $2.4 - 1.6 = 0.8$  MPa. The increase in cytoplasmic pressure is proportional to the molar concentration of glycerol ( $P = imRT$ , where  $P$  is the osmotic pressure of a solution,  $i$

### 6.3 Dynamics of wild type budding yeast osmoregulation pathway

is the Van't Hoff constant,  $m$  is molar concentration,  $R$  is the Avogadro number,  $T$  is the absolute temperature). Hence,  $\Delta P_{cyt} = k_g G$ , where  $k_g$  is a constant, and  $G$  is the molar concentration of glycerol produced in response to activation of the Hog1 pathway. In our model we considered only key components of the pathway. The model is shown in supplementary equation 6.2.

$$\begin{aligned}
\frac{dSho1}{dt} &= B_6 - a_6 Sho1 + \frac{k_6}{(1 + (Hog1/K_6)_6^n)} + \\
&\quad k_{sho1} \frac{((P - k_g * G)/k_p)^2}{(1 + ((P - k_g * G)/k_p)^2)} \\
\frac{dSln1}{dt} &= B_1 - a_1 Sln1 + k_1 \frac{((P - k_g * G)/k_p)^2}{(1 + ((P - k_g * G)/k_p)^2)} \\
\frac{dSsk1}{dt} &= B_2 - a_2 Ssk1 + k_2 \frac{Sln1^{n_1}}{(K_2^{n_1} + Sln1^{n_1})} \\
\frac{dSte11}{dt} &= B_7 - a_7 Ste11 + k_7 \frac{(Sho1/K_7)^{n_7}}{(1 + (Sho1/K_7)^{n_7})} \\
\frac{dPbs2}{dt} &= B_3 - a_3 Pbs2 + k_{31} \frac{(Ssk1/K_{31})^{n_2} + (Ste11/K_{32})^{n_{21}}}{(1 + (Ssk1/K_{31})^{n_2} + (Ste11/K_{32})^{n_{21}})} \\
\frac{dHog1}{dt} &= B_4 - a_4 Hog1 + k_4 \frac{Pbs2^{n_4}}{(1 + (Pbs2/K_4)_4^n + (Ptp2/K_{41})^{n_4})} \\
\frac{dG}{dt} &= -k_{11} G + k_5 \frac{Hog1^{n_{41}}}{(K_5^{n_{41}} + Hog1^{n_{41}})} \\
\frac{dPtp2}{dt} &= B_8 - a_8 Ptp2 + k_8 \frac{Hog1^{n_8}}{(K_8^{n_8} + Hog1^{n_8})}
\end{aligned} \tag{6.2}$$

Parameters of this model were inferred using the simulated annealing based parameter estimation algorithm of the Systems Biology toolbox [Schmidt and Jirstrand \(2006\)](#) to produce similar response as [Krantz et al.](#)



### 6.3 Dynamics of wild type budding yeast osmoregulation pathway

(2004a). The parameter values for the model shown in equation 6.2 are given in table 6.2.

Params	Values	Params	Values	Params	Values	Params	Values	Params	Values
$B_1$	0.001	$k_2$	2.5	$K_6$	1	$n_{21}$	2	$k_5$	1.4
$a_1$	0.9274	$n_1$	2	$k_6$	0.1	$B_4$	0.001	$n_{41}$	2
$P$	0.8	$B_3$	0.001	$n_6$	2	$a_4$	0.5	$B_8$	0.001
$k_1$	2.0233	$a_3$	0.9562	$B_7$	0.001	$K_4$	0.6491	$a_8$	0.2391
$k_g$	0.5	$K_{31}$	0.02	$a_7$	0.5	$K_{41}$	0.9	$K_8$	0.9
$k_p$	0.91	$k_{31}$	1	$K_7$	0.3361	$k_4$	2	$k_8$	0.5
$B_2$	1E-5	$n_2$	2	$k_7$	0.3	$n_4$	2	$n_8$	2
$a_2$	1.655	$B_6$	0	$n_7$	2	$k_{11}$	0.0997	$k_{sho1}$	5
$K_2$	0.9717	$a_6$	1.5	$K_{32}$	0.0192	$K_5$	1.5	—	—

Table 6.2: Parameter values for equation 6.2 which represents the mathematical model of osmotic regulation pathway of wild type yeast cells

In figure 6.6 and 6.3 we show the results of simulation of the model shown in equation 6.2 and the level of glycerol production observed by Krantz et al. (2004a).

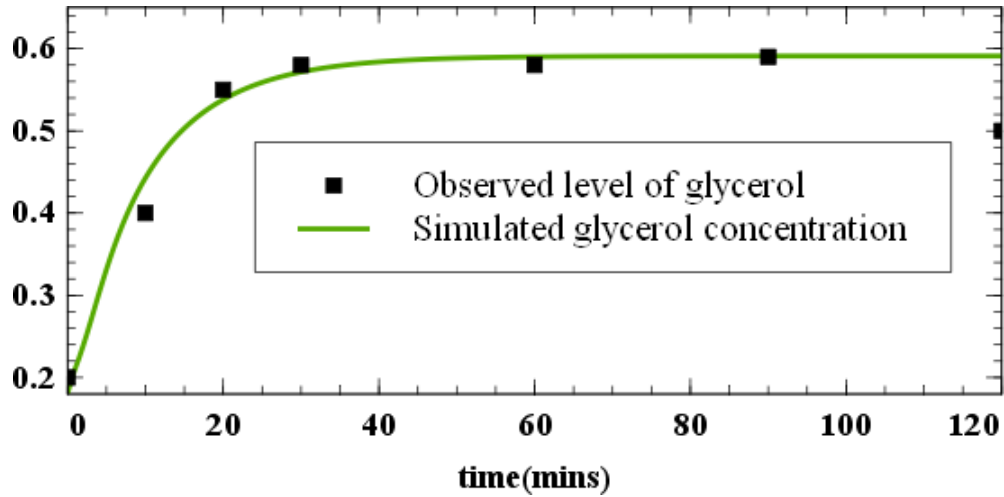
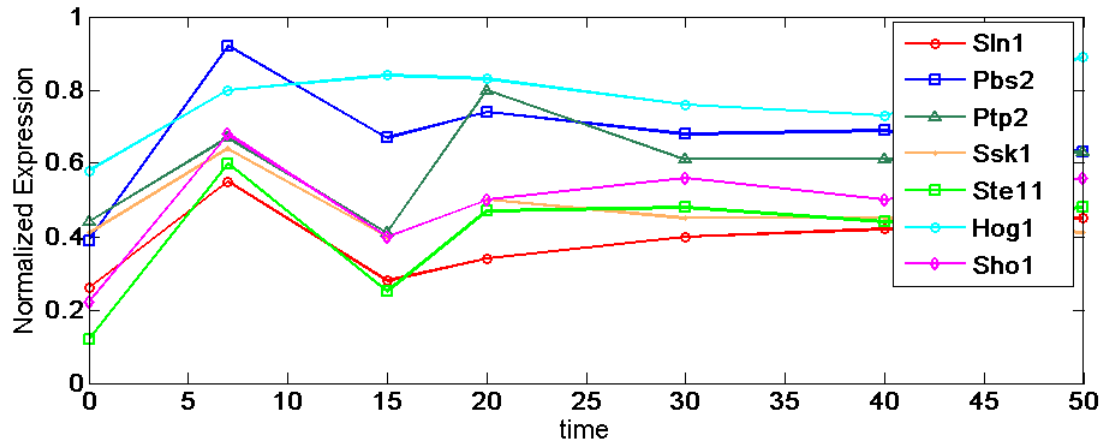
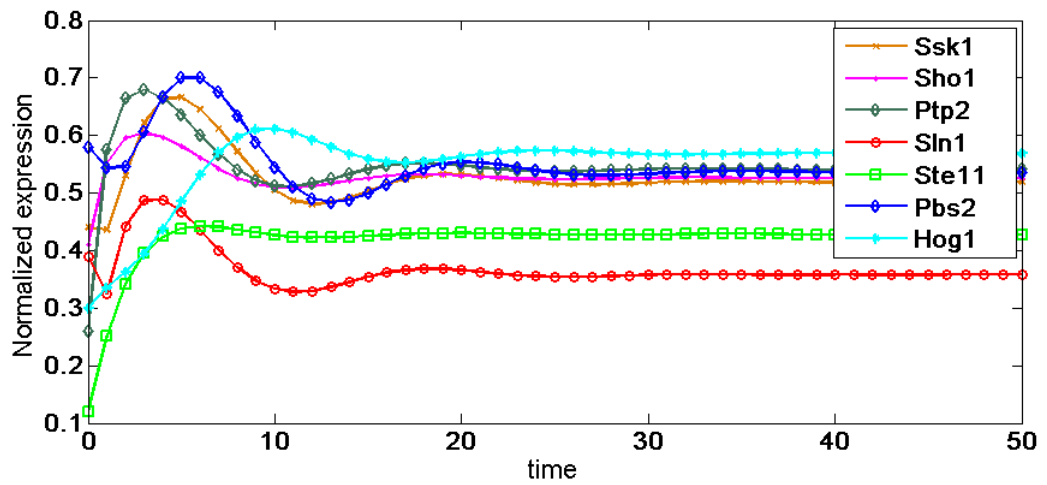


Figure 6.6: The response of the mathematical model of the osmoregulation pathway. The concentration of the end product glycerol of the osmoregulation pathway as observed by Krantz et al. (2004b) is plotted against the glycerol concentration simulated by our model.

### 6.3 Dynamics of wild type budding yeast osmoregulation pathway



(a)



(b)

Figure 6.7: The simulation results of osmoregulation pathway of yeast. (a) shows the concentration levels observed by [Krantz et al. \(2004b\)](#) and (b) shown the simulation results of our model.

Our analysis suggests that the pathway does not exhibit any bifurcation due to stochastic fluctuations in its parameters. Bifurcation diagrams against some of its parameters are shown in figure 6.8. It can be seen from figure 6.8 that the osmotic shock response system of budding yeast is very stable and robust against parametric perturbations, i.e. fluctuations in reaction rates.

#### 6.3.1 The roles of ptp2,3

The osmoregulation of yeast has two parallel signalling pathways, one senses external osmotic pressure via Sln1 and the other via Sho1 protein. Both these pathways consist of long chains of MAPK cascades and finally converge by activating Pbs1. Disrupting one of these MAPK cascades does not produce any detectable phenotype. However, disrupting the pathway after the convergence of two MAPK cascades by mutating Pbs2 or Hog1 produces lethal phenotype during hyperosmotic shock (Sakumoto et al. (2002)). Interestingly, disrupting the small negative feedback mechanisms mediated by Hog1 and Ptp2,3 by mutating Ptp2,3 produces no lethal phenotype either (Sakumoto et al. (2002)). Hence, we presume that the roles of the Ptp2,3 proteins in the functioning of the osmoregulation pathway of budding yeast may be beneficial but not essential. To understand the exact roles of Ptp2,3 in the reaction mechanism of budding yeast to hyper osmotic shock we analysed the dynamics of a  $ptp2,3\Delta$  mutant and compared it with that of a wild type system.

### 6.3 Dynamics of wild type budding yeast osmoregulation pathway

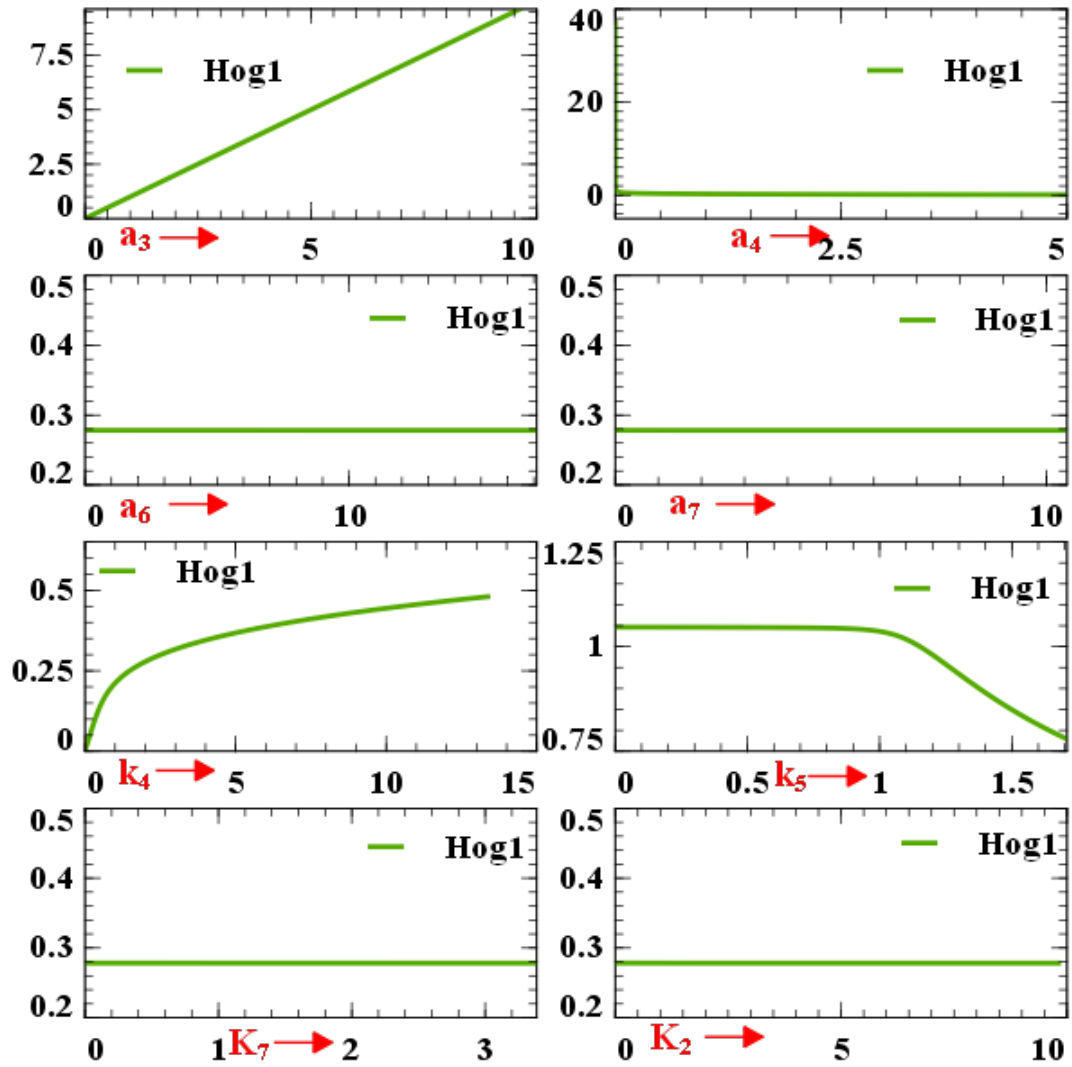


Figure 6.8: Bifurcation analysis of the wild type budding yeast osmoregulation pathway. No bifurcation occurs due to changes in the parameters of the osmoregulation pathway of budding yeast.

### 6.3.2 The dynamics of $\text{ptp2,3}\Delta$ mutant

A  $\text{ptp2,3}\Delta$  mutant is realised by making  $K_{41}$  arbitrarily large, in our case  $K_{41} = 10$ . Bifurcation analysis of the  $\text{ptp2,3}\Delta$  mutant suggests that the osmoregulation pathway of the mutant cell exhibits Hopf bifurcation due to small perturbations in many of its parameters, see figure 6.9. In other words, in a  $\text{ptp2,3}\Delta$  mutant the osmoregulation pathway has rather fragile dynamics and may oscillate due to small changes in its reaction rate constants. Oscillation is undesirable in stress response mechanisms such as this because it results in sluggish and fluctuating response to environmental stress. Since in a wild type cell the osmoregulation pathway does not produce any oscillation we conclude that the Ptp2,3 mediated small feedback mechanism act as a damper module and stabilises possible oscillations due to fluctuations in its reaction rate constants. However, since  $\text{ptp2,3}\Delta$  mutants exhibit no deleterious phenotype we conclude that such a module is beneficial but not essential for the survival of an yeast cell in an episode of hyperosmotic shock.

### 6.3.3 The dominant and auxiliary feedback of yeast osmoregulation pathway

Our analysis suggests that the Sln1 mediated long feedback mechanism acts as a dominant feedback loop and the Ptp2/3 mediated small negative feedback loops act as 'auxiliary' systems. According to our hypothesis

### 6.3 Dynamics of wild type budding yeast osmoregulation pathway

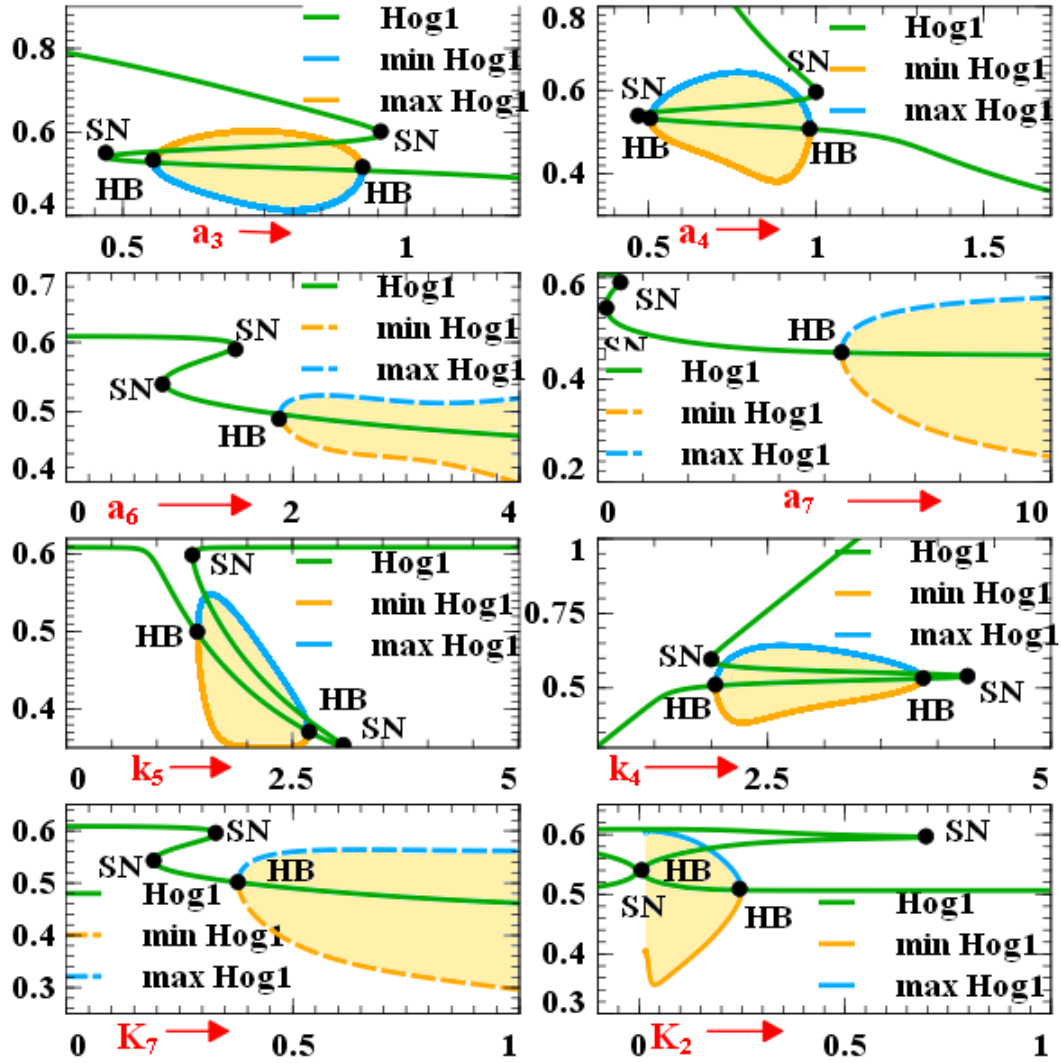


Figure 6.9: Bifurcation diagrams of the *ptp2,3Δ* mutant osmoregulation model. The model shows both saddle node(SN) which promotes bistability and Hopf bifurcation (HB) which promotes oscillation. The oscillatory regions are highlighted in yellow.

the Sln1 mediated long feedback loop should be conserved due to its dominant functional role and the Ptp2/3 mediated small feedback loop may or may not be conserved due its auxiliary role in the functioning of this pathway.

#### 6.3.4 Evolution of osmoregulation pathway in different yeast species

We reconstructed the osmoregulation pathway of fission yeast from literature ([Aiba et al. \(1998\)](#); [Boisnard et al. \(2008\)](#); [Cottarel \(1997\)](#); [Ikner and Shiozaki \(2005\)](#)) evidence and compare its structural conservation with that of budding yeast. For comparison we present the osmoregulation pathways of budding yeast (*S. cerevisiae*), fission yeast (*S. pombe*) and *Candida lusitanae* in figure 6.10. It can be seen from figure 6.10 that the entire osmoregulation pathway including the small feedback loop mediated by Ptp2,3 is conserved in the distantly related yeast species budding and fission yeast. However, in case of *Candida lusitanae*, though the main Hog1 mediated feedback mechanism is conserved there are no evidence of the Ptp2,3 mediated negative feedback regulation of Hog1 ([Boisnard et al. \(2008\)](#)). This presents an excellent evidence of our argument of conservation of the long feedback loops in protein interaction networks due to their biological importance and relative evolutionary plasticity of auxiliary small feedback mechanisms which provides extra functional assistance. Our conclusion is summarised in figure 6.11

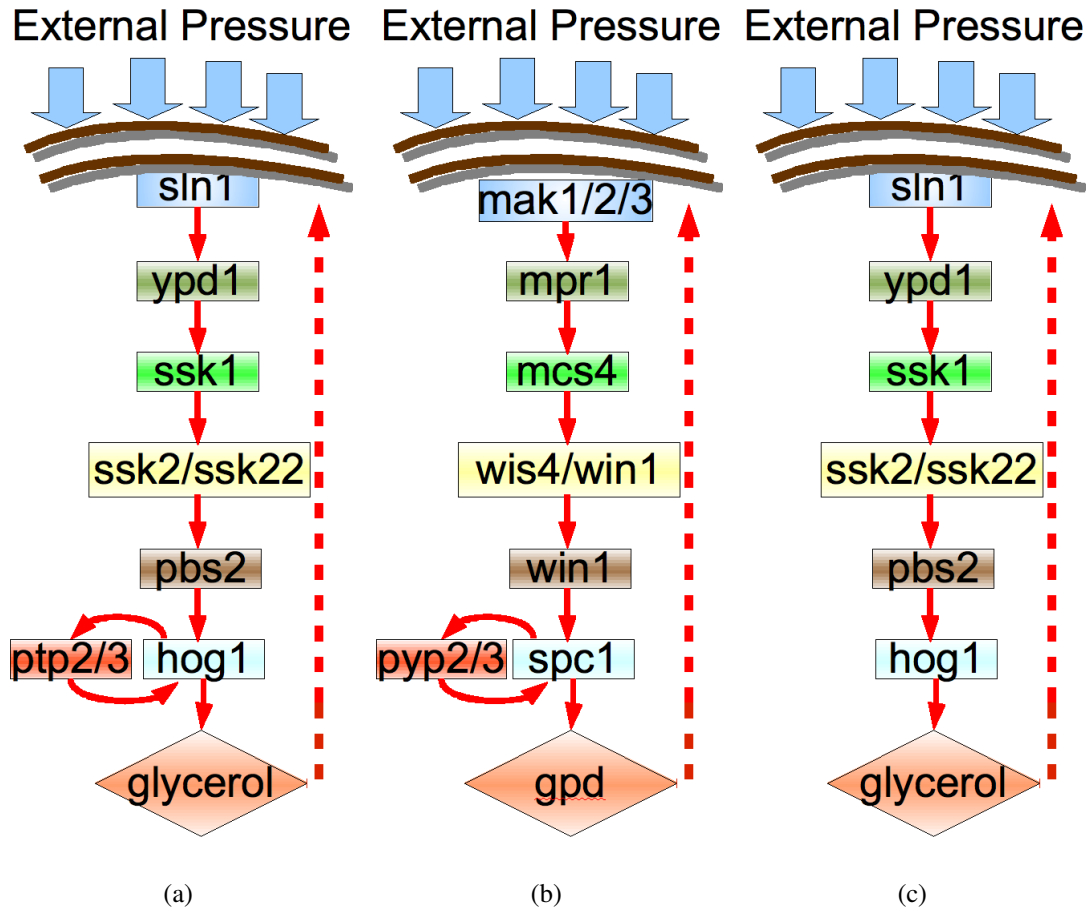


Figure 6.10: Conservation of the osmoregulation pathway. Schematic diagrams of the osmoregulation pathways of budding (a) and fission yeast (b) and *Candida lusitanae* (c). Boxes of same color represent homologous genes, i.e., *mcs4*, *ssk1*; *wak1*, *ssk2,22*; *wis1*, *pbs2*; *sty1*, *hog1* and *pyp2*, *ptp2* are homologous Cottarel (1997). Similarly, *mak1,2,3*, *sln1*; *mpr1*, *ypd1* are homologous Santos and Shiozaki (2001). Some of interactions of the osmoregulation pathways are taken from Aiba et al. (1998); Ikner and Shiozaki (2005). (c) The osmoregulation pathway of *Candida lusitanae* as described by Boissard et al. (2008).





### 6.4 A critical assessment and concluding remarks

Like previous chapter, the main conclusions in this chapter are also results of local bifurcation analysis. The scopes and limitations of this type of analysis are discussed in the previous chapter. However, there are experimental evidence to support the results of bifurcation analysis of the pheromone response pathway [Paliwal et al. \(2007\)](#). On the other hand, the parameters of the osmoregulation pathway model are inferred from time series data published by [Krantz et al. \(2004b\)](#). The data sets consist of eight time series with seven time points each and the number of estimated parameters are 44. High dimensionality of the parameter space results in inaccurate parameter estimations especially in the presence of small number of coarse time series data. Though the point estimates in this study seem to fit the observed data reasonably well, a better estimate can be achieved by using better data and models with smaller set of parameters. However, similar dynamic properties of the same pathway was also found by [Hersen et al. \(2007\)](#).

In most cases feedback mechanisms in yeast biochemical networks exhibit monostability, bistability, multistability, graded response or oscillatory dynamics. In almost all cases the system converges to a global homeostasis depending on their initial component concentrations and reaction rate constants. Oscillatory systems are exceptions. The kinetic importance of oscillatory modules and their implications in the evolution of bio-molecular circuits are discussed in the next chapter.

## Chapter 7

### Evolution of feedback modules which generate autonomous oscillation

Oscillation arises from negative feedbacks in biochemical interaction networks. Several modules with feedback mechanisms in yeast biochemical networks such as the cell cycle module, the ultradian clock etc are found to produce sustained autonomous oscillation. The cell cycle module periodically divide cells [Chen et al. \(2000\)](#) whereas the ultradian clock serves as a cellular clock which synchronises different biological processes [Tu et al. \(2005b\)](#). The dynamics of oscillating modules are different from any other stable dynamic system in the sense that oscillating modules never reaches homeostasis. Additionally, when a cyclic module oscillates in its normal operating condition it provides oscillating input to its downstream gene or protein sub-networks. [Liu et al. \(2005\)](#) suggested that, in this case the unforced equilibrium of the downstream

---

sub-network collapses and it exhibits various periodic and aperiodic oscillations depending on the reaction rates and the period and amplitude of the input oscillation. In other words the subnetwork downstream to an oscillating module shows forced oscillation.

For example, [Spellman et al. \(1998\)](#) found that around 800 yeast genes oscillate in synchrony with its cell cycle. To find out whether these genes are indeed regulated by the yeast cell cycle oscillator we analysed the hierarchy of yeast GRN from the transcription regulation data published by [Lee et al. \(2002\)](#); [Luscombe et al. \(2004a\)](#); [Milo et al. \(2002\)](#). We find that most (approximately 66%) of the genes which were found to oscillate with the same period as cell cycle (see [Spellman et al. \(1998\)](#) for a comprehensive list) are indeed directly or indirectly regulated by the genes which generate the oscillation ([Lipman et al. \(2002\)](#)). However, theoretically we should have found all the genes which oscillate with the same period as cell cycle lie downstream to the the genes which generates cell cycle oscillation in the first place. There could be several reasons behind the fact that we found only 66% of the genes which oscillate in synchrony with cell cycle lie downstream to the cell cycle genes. Two main reasons are as follows: Firstly, [Spellman et al. \(1998\)](#)'s method is based on a correlation measure of Fourier coefficients of the time series of mRNA concentrations. Such method depends on the quality of mRNA time series data and the threshold chosen for the correlation score. A higher correlation score improves the percentage as we find that 86% of the 200 genes with the highest correlation score lie downstream to the cell cycle genes in the gene regulatory network (GRN) of yeast. The noise in the mRNA time series data also affects the analysis. Secondly,

---

the GRN data we used for our analysis is noisy too which might affect our estimate.

However, from the above analysis it can be claimed unambiguously that yeast cell cycle module affects the dynamics of a large number of genes. Because of such diverse kinetic influences of oscillating modules on their downstream subnetworks any change in these modules may result in wide spread effects on biological processes which are orchestrated in synchrony with them. For example, biological processes such as DNA replication, microtubule cytoskeleton organization, organelle fission, nuclear division, chromatid segregation and different metabolic processes are governed by genes and proteins which are found to oscillate in synchrony with the cell cycle ([Spellman et al. \(1998\)](#)). Disrupting a cell cycle oscillator by mutating its components produces a wide range of symptoms, e.g., delayed cell division, prevention of DNA replication and even cell death ([Chen et al. \(2000\)](#)). Hence, we suggest that these modules might be conserved because of their dynamic influences on large parts of an organism's GRNs.

Although the major transcription factors related to cell cycle network and their DNA binding sites are conserved in both budding and fission yeast and to a varying extent also in Metazoa, there are clear differences in the cell cycle network circuitry reflecting evolutionary rewiring [Bahler \(2005\)](#). Such rewiring adjusts the differences in inter-species cell cycle phasing [Bahler \(2005\)](#). A core mechanism that generates cell cycle oscillations are the oscillating expression levels of cyclins, the proteins which regulate the catalytic activities of CDKs, the kinases that drive

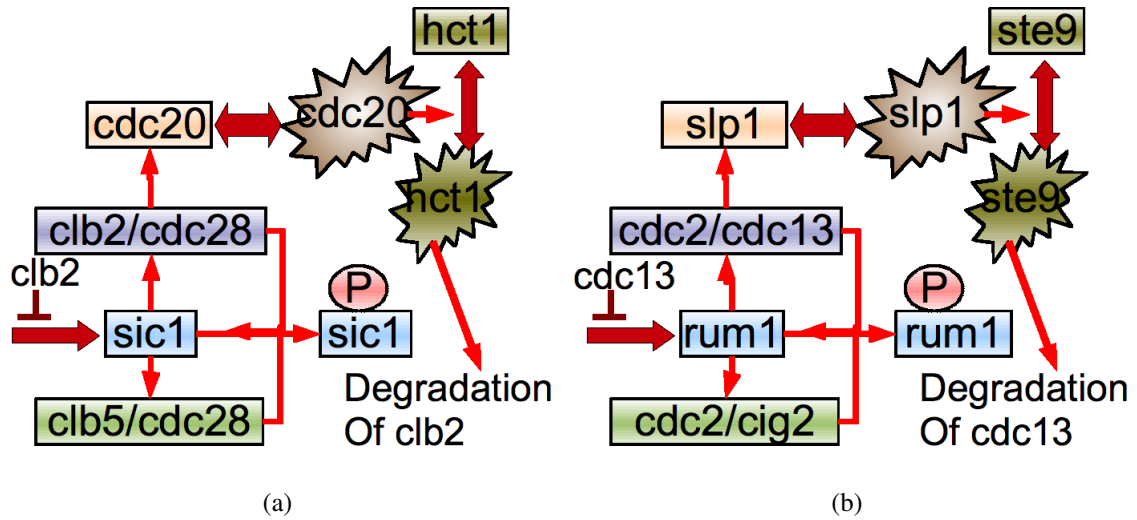


Figure 7.1: Conservation of a prototypical cell cycle oscillator in budding (a) and fission (b) yeast. The interactions of the budding yeast network are from [Chen et al. \(2000\)](#), and of the fission yeast network from [Sveiczer et al. \(2004\)](#). Boxes of the same colour represent homologous genes. The prototypical yeast cell cycle module is conserved in both yeasts. A rather simplistic description of the prototypical cell cycle oscillator is as follows. Cyclins, Clb2, Clb5(Cdc13,Cig2) are inhibited by Sic1(Rum1) at the beginning(G1 phase) of *S. cerevisiae* (*S. pombe*) cell cycle. As the cell grows Cln2/3(not shown in the figure, similar mechanism in *S.pombe* is not yet clear) degrades Sic1(Rum1) and release Clb2(Cdc13) which represses Sic1(Rum1) thus letting free Clb2, Clb5(Cig2,Cdc13) accumulate in the cell which triggers the G1-S transition. In middle and late phases of the cell cycle, Clb2(Cdc13) activates Cdc20(Slp1), which activates Hct1(Ste9). Activated Hct1(Ste9) degrades Clb2(Cdc13) releasing Sic1 from its inhibitory effect and the cell returns to G1 resembling an oscillator.

cell cycle progression. To locate such prototypical oscillators we looked for conserved negative feedback mechanisms involving cyclins in both budding and fission yeast. We find that such a minimal oscillator indeed exists and is conserved in both yeast species (Fig. 7.1(a), 7.1(b)). The genes involving these prototypes have considerable sequence similarity and functional homology.

## Chapter 8

### Conclusion

Evolution of GRNs has been mainly studied in the context of body plan development in model organisms such as arthropods ([Damen, 2007](#)), nematodes ([Ge et al., 2006](#)), and echinoderms ([Hinman et al., 2009](#)). From these studies it is clear that GRNs specify the evolution of body plans. Surprisingly, it also became apparent that the network architecture seems to be conserved better than the components exerting specific functions, and that such network modules tend to be organized as evolutionary conserved "kernels" ([Hinman et al., 2009](#)). This, in turn, suggests that network architecture may determine the evolvability of different parts of the network.

We tested this hypothesis using a combined topological and dynamic analysis of yeast biochemical networks. The results provide evidence to

---

support this hypothesis. They show that the acyclic parts of a biochemical interaction network evolve more rapidly than the cyclic parts, which are relatively rigid against such changes. Furthermore, we show that long range and oscillatory feedback mechanisms are more likely to resist evolutionary changes. However, different oscillating modules in yeast have appeared at many different stages of evolution. For example, the most basic and vital oscillator, i.e. the prototypical cell cycle oscillator, is found conserved in even the most distantly related yeast species such as *S. cerevisiae* and *S. pombe*. On the other hand, the metabolic oscillators are rather species specific and thought to have appeared after divergence of the *Saccharomycetaceae* family from the *S. cerevisiae* lineage.

Our results also suggest that acyclic parts of networks are functionally much more robust toward structural changes and hence can tolerate evolutionary changes. This reveals an unexpected link between network design features and evolution. It also shows that there is an intrinsic balance between robustness of design and potential for evolution. An interesting future question is to define why and under what conditions acyclic parts become cyclic and evolutionary engrained.

Our study provides some interesting insights into the evolution of biochemical networks. Most of our predictions are based on theoretical analysis and we tested our predictions by carrying out both statistical analysis of large metabolic networks and case studies of individual biochemical modules of yeast species. The best way to validate our conclusions is to carry out systematic statistical analysis of biochemical networks of different species. But the main obstacle in doing so is the



---

lack of interaction data for most of the species. Though the genomes of many yeast species are sequenced, their interactomes are yet to be determined. Another big obstacle is in the concept of the statistical analysis itself. In order to determine whether the majority of the feedback loops are conserved or not, we have to accurately determine the feedback loops present in each of biochemical networks being compared. As we already discussed before, determining feedback loops in a biochemical network needs precise knowledge about the network itself. Such information is not available for even the most well studied organism such as *S. cerevisiae*. The interaction data used in such studies often contain false negatives and false positives which can generate false negative and false positive feedback loops. For accurate reliable inference of structural properties it is also necessary to develop better strategies to deal with noise present in interaction data.

Even if, in a hypothetical scenario, precise interaction data were available for many species, determining the structural properties of these networks would have been difficult because a large proportion of these interaction data are binary protein interaction data. It is difficult to estimate the direction of signal flow in biochemical networks made of such interactions. In our study, we have used an algorithm developed by [Liu et al. \(2009\)](#) to determine directionality (partially) in protein interaction networks. [Liu et al. \(2009\)](#)'s method is one of the first of its kind and yet to be approved by the larger computational biology community. At present such methods can only be used to gather rough estimates of certain structural properties.

---

Hence, designing statistical analysis to compare the biochemical networks of different yeast species in a systematic way is not possible in the current context. However such an analysis will be possible when more data and sophisticated computational methods are available to deal with the data in a more efficient way. Hence, a future direction of this study will be to develop databases of biochemical reactions for non metabolic pathways of multiple yeast species. Though such databases exist for metabolic pathways, a collective database of biochemical reactions which take part in non metabolic pathways is not yet present. Developing such database involves extensive literature curation and development of new techniques, such as [Saleem et al. \(2010\)](#), to gather data accurately in a genomic scale. Once reliable reactomes are available for more than one yeast species we shall be able to carry out statistical analysis to determine evolutionary conservation in different parts of biochemical networks.

## References

- C. Adami. Evolution: Reducible complexity. *Science*, 312:61–63, 2007. [49](#), [52](#)
- I. Agrafioti et al. Comparative analysis of *Saccharomyces cerevisiae* and *Caenorhabditis elegans* protein interaction networks. *BMC Evol. Biol.*, 5(23), 2005. [2](#)
- H. Aiba et al. Isolation and characterization of high-osmolarity-sensitive mutants of fission yeast. *J. Bacteriology*, 180:5038–5043, 1998. [xxv](#), [xxvi](#), [170](#), [171](#), [172](#)
- S. Aligianni, D. H. Lackner, S. Klier, G. Rustici, B. T. Wilhelm, S. Marguerat, S. Codlin, A. Brazma, R. A. M. Bruin, and J. Bahler. The fission yeast homeodomain protein Yox1p binds to MBF and confines MBF-dependent cell-cycle transcription to G1-S via negative feedback. *PLOS Genetics*, 5:e1000626, 2009. [100](#)
- A. Anders, H. Lilie, K. Franke, L. Kapp, J. Stelling, E. D. Gilles, and K. D. Breunig. The galactose switch in *Kluyveromyces lactis* depends on nuclear competition between GAL4 and GAL1 for GAL80 binding. *Journal of Biological Chemistry*, 281:29337–29348, 2006. [101](#)

## REFERENCES

---

- M. Ashburner et al. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000. [61](#)
- S. Asur, D. Ucar, and S. Parthasarathy. An ensemble framework for clustering proteinprotein interaction networks. *Bioinformatics*, 23: i29–i40, 2007. [69](#)
- G. D. Bader et al. BIND—the Biomolecular Interaction Network Database. *Nucleic Acids Research*, 29(1):242–245, 2001. [58](#), [60](#), [66](#)
- J. Bahler. Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet.*, 39:69–94, 2005. [176](#)
- N. N. Batada et al. Stratus not altocumulus: a new view of the yeast protein interaction network. *PLos Biology*, 4:e317, 2006. [67](#)
- S. Betina, P. Goffrini, I. Ferrero, and M. W. Louvel. RAG4 gene encodes a glucose sensor in *Kluyveromyces lactis*. *Genetics*, 158:541–548, 2001. [103](#)
- N. Bharhadwaj and H. Lu. Correlation between gene expression profiles and protein protein interactions within and accross genome. *Bioinformatics*, 21(11):2730–2738, 2005. [56](#)
- P. J. Bhat. *Galactose Regulon of Yeast from Genetics to Systems Biology*. Springer, 2008. [101](#)
- J. D. Bloom and C. Adami. Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein protein interaction data. *BMC Evol. Biol.*, 3(21), 2003. [3](#)

## REFERENCES

---

- E. M. Boczko et al. Structure theorems and the dynamics of nitrogen catabolite repression in yeast. *PNAS*, 102:5647–5652, 2005. [xvii](#), [105](#), [109](#), [111](#), [113](#), [115](#), [116](#), [117](#), [118](#)
- S. Boissard et al. Insight into the role of hog pathway components Ssk2p, Pbs2p and Hog1p in the opportunistic yeast *Candida lusitanae*. *Eucaryotic cell*, 7(12):2179–2183, 2008. [xxv](#), [xxvi](#), [170](#), [171](#), [172](#)
- G. Butler et al. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature*, 459(657-662), 2009. [2](#)
- K.C. Chen et al. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell*, 11:369–391, 2000. [xxvii](#), [105](#), [174](#), [176](#), [177](#)
- S. Ciliberti, O. C. Martin, and A. Wagner. Innovation and robustness in complex regulatory gene networks. *PNAS*, 104:13591–13596, 2007. [5](#)
- A. Clements, D. Bursac, et al. The reducible complexity of a mitochondrial molecular machine. *PNAS*, doi10.1073pnas.0908264106, 2009. [48](#), [52](#)
- J. A. Coffman et al. Cross regulation of four GATA factors that control nitrogen catabolic gene expression in *saccharomyces cerevisiae*. *Journal of Bacteriology*, 179:3416–3429, 1997. [120](#)
- S. Cooper. Distinguishing between linear and exponential cell growth during the division cycle: Single-cell studies, cell-culture studies, and the object of cell-cycle research. *Theoretical Biology and Medical Modelling*, 3:doi:10.1186/1742–4682–3–10, 2006. [32](#)

## REFERENCES

---

- G. Cottarel. Mcs4, a two-component system response regulator homologue, regulates the *Schizosaccharomyces pombe* cell cycle control. *Genetics*, 147:1043–1051, 1997. [xxv](#), [xxvi](#), [170](#), [171](#), [172](#)
- W. Damen. Evolutionary conservation and divergence of the segmentation process in arthropods. *Developmental Dynamics*, 236:1379–1391, 2007. [10](#), [178](#)
- E. H. Davidson and D.H. Erwin. Gene regulatory networks and the evolution of animal body plans. *Science*, 311:796–800, 2006. [4](#), [5](#), [10](#)
- M. K. Davidson, H. K. Shandilya, K. Hirota, K. Ohta, and Wahls P. W. Atf1-Pcr1-M26 complex links stress-activated MAPK and cAMP-dependent protein kinase pathways via chromatin remodeling of Cgs2. *Journal of Biological Chemistry*, 279:50857–50863, 2004. [98](#)
- S.J. Dixon et al. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *PNAS*, 105:16653–16658, 2008. [3](#), [79](#), [81](#)
- M. Feinberg. Complex balancing in general kinetic systems. *Arch. Rational. Mech. Anal.*, 49:187–194, 1972. [34](#), [35](#), [37](#), [39](#), [40](#), [41](#), [46](#), [49](#)
- K. M. Flick et al. Grr1 dependent interactivation of Mth1 mediate glucose induced dissociation of Rgt1 from HXT gene promoters. *Mol. Biol. Cell*, 14(8):3230–3241, 2003. [135](#)
- H. B. Fraser, D. P. Wall, and A. E. Hirsh. A simple dependence between protein evolution rate and the number of protein protein interactions. *BMC Evol. Biol.*, 3(11), 2003. [2](#), [3](#)

## REFERENCES

---

- A. C. Gavin et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440:631–636, 2006. [66](#)
- H. Ge, C. M. Player, and L. Zou. Toward a global picture of development : Lessons from genome-scale analysis in *Caenorhabditis elegans* embryonic development. *Developmental Dynamics*, 235:2009–2017, 2006. [10](#), [178](#)
- P. Gennemark, B. Nordlander, S. Hohmann, and D. Wedelin. A simple mathematical model of adaptation to high osmolarity in yeast. *In Silico Biology*, 6(3):193–214, 2006. [160](#)
- P. Gervais, P. Molin, A. Marechal, and C. Herail-Fossereau. Thermodynamics of yeast cell osmoregulation: Passive mechanisms. *Journal of Biological Physics*, 22:73–86, 1996. [162](#)
- M. Girolami and B. Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B*, 73, 2011. [142](#)
- D. Graur and W. H. Li. *Fundamentals of molecular evolution*. Sinaur Associates, Massachussets, 2000. [1](#)
- R. Grima and S. Schnell. A systematic investigation of the rate laws valid in intracellular environment. *Biophysical Chemistry*, 124(1): 1–10, 2006. [20](#)
- J. Gunawardena. Chemical reaction network theory for in-silico biologists. *Lecture notes*, 2003. [37](#), [46](#)
- M. W. Hahn, G. C. Conant, and A. Wagner. Molecular evolution in

## REFERENCES

---

- large genetic networks: does connectivity equal constraint? *Journal of Molecular Evolution*, 58:203–211, 2004. [2](#)
- L. Hakes, W. Pinney, J. D.L. Robertson, and S.C. Lovell. Protein-protein interaction networks and biologywhat’s the connection. *Nature Biotechnology*, 26:69–72, 2008. [67](#)
- C. Harrison, S. Katayama, S. Dhut, D. Chen, N. Jones, J. Bahler, and T. Toda. SCF-Pof1-ubiquitin and its target Zip1 transcription factor mediate cadmium response in fission yeast. *The EMBO journal*, 24: 599–610, 2005. [99](#)
- M. J. Herrgard et al. A consensus yeast metabolic reconstruction obtained from a community approach to systems biology. *Nature Biotechnology*, 26:1155–1160, 2008. [145](#)
- P. Hersen, M. N. McClean, L. Mahadevan, and S. Ramanathan. Signal processing by the hog map kinase pathway. *PNAS*, 105:7165–7170, 2007. [173](#)
- I. Herskowitz. Map kinase pathways in yeast: For mating and more. *Cell*, 80:187–197, 1995. [xxiv](#), [157](#), [158](#)
- V. F. Hinman, K. A. Yankuraa, and B. S. McCauleya. Evolution of gene regulatory network architectures: Examples of subcircuit conservation and plasticity between classes of echinoderms. *Biochimica et Biophysica Acta*, 1789:326–332, 2009. [10](#), [107](#), [178](#)
- C. E. Horak, N. M. Luscombe, J. Qian, P. Bertone, S. Piccirillo, M. Gerstein, and M. Snyder. Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes and Development*, 16: 3017–3033, 2002. [100](#)



## REFERENCES

---

- F. Horn and R. Jackson. General mass action kinetics. *Arch. Rational. Mech. Anal.*, 47:81–116, 1972. [34](#), [35](#), [37](#), [46](#), [49](#)
- C.H. Huang, I.F. Fang, J.J.P. Tsai, and Ng. K.L. Topological robustness of the protein protein interaction networks. *RECOMB*, LNBI-4023: 166–177, 2007. [60](#), [65](#)
- D.A. Hughes, N. Yabana, and M. Yamamoto. Transcriptional regulation of a ras nucleotide-exchange factor gene by extracellular signals in fission yeast. *Journal of Cell Science*, 107:3536–3542, 1994. [xxiv](#), [157](#), [158](#)
- A. Ikner and K. Shiozaki. Yeast signalling pathways in the oxidative stress response. *Mutation Research*, 569:13–27, 2005. [xxv](#), [xxvi](#), [170](#), [171](#), [172](#)
- Y. Imai and M. Yamamoto. *Schizosaccharomyces pombe* Sxa1+ and Sxa2+ encode putative proteases involved in the mating response. *Molecular and Cellular Biology*, 12:1827–1834, 1992. [xxiv](#), [157](#), [158](#)
- S. Jeffrey and V. Brown. Glucose sensing network in *candida albicans*: a sweet spot for fungal morphogenesis. *Eukaryotic Cell*, 8(9):1314–1320, 2009. [xxii](#), [135](#), [140](#), [141](#)
- W. Jonkers and M. Rep. Lessons from fungal f-box proteins. *Eukaryotic Cell*, 8:677–695, 2009. [99](#)
- I. K. Jordan, Y. I. Wolf, and Koonin E. V. No simple dependence between protein evolution rate and the number of protein protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol. Biol.*, 3(1), 2003. [2](#)

## REFERENCES

---

- M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and T. Yamanishi. Kegg for linking genomes to life and the environment. *Nucleic Acids Research*, 36:D480–D484, 2008. [87](#), [95](#), [104](#), [105](#)
- A. Kaniak, X. Zhixiong, D. Macool, J. H. Kim, and M. Johnston. Regulatory network connecting two glucose signal transduction pathways in *saccharomyces cerevisiae*. *Eukaryotic Cell*, 3:221–231, 2004. [xx](#), [105](#), [109](#), [127](#), [129](#), [130](#), [132](#)
- N. Kashtan and U. Alon. Spontaneous evolution of modularity and network motifs. *PNAS*, 102:13773–13778, 2005. [5](#)
- M. Kellis, N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, 423:241–245, 2003. [2](#)
- J. R. Kim, Y. Yoon, and Cho K. H. Coupled feedback loops form dynamic motifs in cellular network. *Biophysical Journal*, 94:359–365, 2007. [18](#)
- E. Klipp, B. Nordlander, R. Kruger, P. Gennemark, and S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nature Biotechnology*, 23:975–982, 2005. [105](#), [109](#), [146](#), [160](#)
- B. Kofahl and E. Klipp. Modelling the dynamics of the yeast pheromone pathway. *Yeast*, 21:831–850, 2004. [105](#), [147](#), [148](#), [150](#), [151](#), [152](#)
- R. Kopelman. Fractal reaction kinetics. *Science*, 241:1620–1626, 1988. [20](#)

## REFERENCES

---

- D. Kornitzer. Monitoring protein degradation. *Methods Enzymol.*, 351: 639–47, 2002. [32](#)
- M. Krantz et al. Anaerobicity prepares *Saccharomyces cerevisiae* cells for faster adaptation to osmotic shock. *Eukaryotic Cell*, 3:1381–1390, 2004a. [162](#), [163](#), [164](#)
- M. Krantz et al. Anaerobicity prepares *Saccharomyces cerevisiae* cells for faster adaptation to osmotic shock. *Eukaryotic Cell*, 3:1381–1390, 2004b. [xxiv](#), [164](#), [165](#), [173](#)
- A. Krishnan, M. Tomita, and A. Guilian. Evolution of gene regulatory networks: robustness as an emergent property. *Physica A: Statistical Mechanics and its applications*, 387:2170–2186, 2008. [5](#)
- A. Kudlicki, M. Rowicka, and Z. Otwinowski. Sceptrans: an online tool for analyzing periodic transcription in yeast. *Bioinformatics*, 12: 1559–1561, 2007. [xx](#), [133](#)
- Y. K. Kwon and K. H. Cho. Analysis of feedback loops and robustness in network evolution based on boolean models. *BMC Bioinformatics*, 8(1):430, 2007. [4](#)
- Y. K. Kwon and K. H. Cho. Coherent coupling of feedback loops: a design principle of cell signaling networks. *BMC Bioinformatics*, 24: 1926–32, 2008. [109](#), [110](#)
- J. Lakshmanan, A. L. Mosley, and S. Ozcan. Repression of transcription by Rgt1 in the absense of glucose requires Std1 and Mth1. *Current Genetics*, 44:19–25, 2003. [135](#)

## REFERENCES

---

- T. I. Lee et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298(5594):700–804, October 2002. [xvi](#), [105](#), [108](#), [175](#)
- Z. Liang, M. Xu, M. Teng, and L. Niu. Comparison of protein interaction networks reveals species conservation and divergence. *BMC Bioinformatics*, 7:457, 2006. [80](#), [81](#)
- D. J. Lipman, A. Souvorov, E. V. Koonin, A. R. Panchenko, and T. A. Tatusova. The relationship of protein conservation and sequence length. *BMC Evol. Biol.*, 2(20 doi:10.1186/1471-2148-2-20), 2002. [2](#), [175](#)
- J. Liu, J. W. Crawford, and K. I. Leontiou. Collapse of single stable states via a fractal attraction basin: analysis of a representative metabolic network. *Proceedings of the Royal Society A*, 461:2327–2338, 2005. [174](#)
- R. Liu and H. Ochman. Stepwise formation of the bacterial flagellar system. *PNAS*, 104:7116–7121, 2007. [48](#), [52](#)
- W. Liu et al. Proteome-wide prediction of signal flow direction in protein interaction networks based on interacting domains. *Molecular and Cellular Proteomics*, 8:2063–2070, 2009. [x](#), [57](#), [58](#), [63](#), [64](#), [68](#), [69](#), [180](#)
- J. A. Low and W. R. Atchly. Molecular evolution of the GATA family of transcription factors: conservation within the DNA binding domain. *Journal of Molecular Evolution*, 50(2):103–115, 2000. [xix](#), [126](#), [128](#)
- N. M. Luscombe, M. M. Babu, H. Yu, M. Snyder, S.A. Teichmann,

## REFERENCES

---

- and M. Gerstein. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431:308–312, 2004a. [175](#)
- N. M. Luscombe, M. M. Babu, H. Yu, M. Snyder, S.A. Teichmann, and M. Gerstein. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431:308–312, 2004b. [x](#), [xvi](#), [55](#), [56](#), [57](#), [58](#), [61](#), [62](#), [64](#), [105](#), [108](#)
- S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *PNAS*, pages 11980–11985, 2003. [12](#), [14](#)
- S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon. The incoherent feed-forward loop accelerates the response-time of the GAL system of *Escherichia coli*. *JMB*, 356:1073–1081, 2006. [11](#)
- S. Marcus et al. Shk1, a homolog of the *Saccharomyces cerevisiae* Ste20 and mammalian p65PAK protein kinases, is a component of a Ras/Cdc42 signaling module in the fission yeast *Schizosaccharomyces pombe*. *PNAS*, 92:6180–6184, 1995. [xxiv](#), [157](#), [158](#)
- A. Mazurie, S. Bottani, and M. Vergassola. An evolutionary and functional assessment of regulatory network motifs. *Genome Biology*, 6(4):R35, 2005. [5](#)
- R. Milo et al. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002. [4](#), [5](#), [11](#), [175](#)
- J. Nakabayashi and A. Sasaki. Optimal phosphorylation step number of intracellular signal transduction pathway. *Journal of Theoretical Biology*, 233:413–421, 2005. [159](#)

## REFERENCES

---

- A.N. Nguyen and K. Shiozaki. Heat shock-induced activation of stress map kinase is regulated by threonine and tyrosine specific phosphatases. *Genes and Development*, 13:1653–1663, 1999. [98](#)
- S. Paliwal et al. MAPK-mediated bimodal gene expression and adaptive gradient sensing in yeast. *Nature*, 446:46–51, 2007. [xxiii](#), [105](#), [109](#), [145](#), [147](#), [148](#), [150](#), [151](#), [152](#), [153](#), [154](#), [155](#), [173](#)
- X. Peng et al. Identification of cell cycle-regulated genes in fission yeast. *Mol. Biol. Cell*, 16:1026–1042, 2005. [2](#), [3](#), [77](#), [78](#), [79](#)
- F. Posas, M. Takekawa, and H. Saito. Signal transduction by MAP kinase cascades in budding yeast. *Current Opinion in Microbiology*, 1 (2):175–182, 1998. [145](#)
- M. Proft, F. D. Gibbons, M. Copeland, F. P. Roth, and K. Struhl. Genomewide identification of Sko1 target promoters reveals a regulatory network that operates in response to osmotic stress in *Saccharomyces cerevisiae*. *Eukaryotic Cell*, 4:1343–3152, 2005. [97](#), [98](#)
- N. Pržulj, D. G. Corneil, and I Jurisica. Modeling interactome: Scale free or geometric. *Bioinformatics*, 20(18):3508–3515, 2004. [4](#)
- N. Pržulj, D. G. Corneil, and I Jurisica. Efficient estimation of graphlet frequency distributions in protein protein interaction network. *Bioinformatics*, 22(8):974–980, 2006. [4](#)
- S. A. Raithatha and D. T. Stuart. Meiosis-specific regulation of the *Saccharomyces cerevisiae* S-phase Cyclin CLB5 is dependent on MluI Cell Cycle box (MCB) elements in its promoter but is independent of MCB-binding factor activity. *Genetics*, 169:1329–1342, 2005. [100](#)

## REFERENCES

---

- S. A. Ramsey et al. Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nature Genetics*, 38:1082–1087, 2006. [105](#)
- H. Reinke and D. Gatfield. Genome wide oscillation of transcription in yeast. *Trends in Biochemical Sciences*, 31(4):189–191, 2006. [135](#)
- S. Rogers et al. Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. *Bioinformatics*, 24(24):2894–2900, 2008. [56](#)
- S. J. Ross, V. J. Findlay, and P. Malakasi. Thioredoxin peroxidase is required for the transcriptional response to oxidative stress in budding yeast. *Mol. Biol. Cell*, 11:2631–2642., 2000. [105](#)
- C.A. Rossa and G. Peter. *Biodiversity and Ecophysiology of Yeast*, volume 1. Springer, 2005. [xix](#), [126](#), [128](#)
- G. Rustici et al. Global transcription responses of fission and budding yeast to changes in copper and iron levels: a comparative study. *Genome Biology*, 8(3):R73, 2007. [3](#), [79](#)
- N. Sakumoto et al. A series of double disruptants for protein phosphatase genes in *Saccharomyces cerevisiae* and their phenotypic analysis. *Yeast*, 19(7):587–599, 2002. [166](#)
- R. A. Saleem et al. Integrated phosphoproteomics analysis of a signaling network governing nutrient response and peroxisome induction. *Molecular and Cellular Proteomics*, 9:2076–2088, 2010. [181](#)
- H. Salin, V. Fardeau, E. Piccini, G. Lelandais, V. Tanty, S. Lemoine, C. Jacq, and F. Devaux. Structure and properties of transcriptional

- networks driving selenite stress response in yeasts. *BMC Genomics*, 9:doi:10.1186/1471-2164-9-333, 2008. [102](#)
- J. L. Santos and K. Shiozaki. Fungal histidine kinases. *Sci. STKE*, 98:re1, 2001. [xxv](#), [xxvi](#), [171](#), [172](#)
- M. A. Savageau. Michaelis Menten mechanism reconsidered: Implications of fractal kinetics. *Journal of Theoretical Biology*, 176:115–124, 1995. [20](#)
- H. Schmidt and M. Jirstrand. Systems biology toolbox for matlab: a computational platform for research in systems biology. *Bioinformatics*, 22(514-515), 2006. [113](#), [142](#), [163](#)
- B. Schwikowski, P. Uetz, and S. Fields. A network of protein protein interactions in yeast. *Nature Biotechnology*, 18:1257–1261, 2000. [55](#)
- D. Siegel and Y.F. Chen. The S-C-L graph in chemical kinetics. *Rocky Mountain Journal of Mathematics*, 25:479–489, 1995. [40](#), [42](#), [43](#)
- D. Siegel and M.D. Johnston. Global stability of complex balanced system. *arXiv:1008.1622v1*, 2003. [34](#)
- P.T. Spellman et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, 9:3273–3297, 1998. [175](#), [176](#)
- M. A. Steiger and R. Parker. Analyzing mRNA decay in *Saccharomyces cerevisiae*. *Methods Enzymol.*, 351:648–660, 2002. [32](#)
- S. Suthram, T. Shlomi, E. Ruppin, R. Sharan, and T. Ideker. A direct comparison of protein interaction confidence assignment schemes. *BMC Bioinformatics*, 7:360, 2006. [67](#)



## REFERENCES

---

- A. Sveiczer, J.J. Tyson, and B. Novak. Modelling the fission yeast cell cycle. *Briefings in Functional Genomics and Proteomics*, 2:298–307, 2004. [xxvii](#), [177](#)
- D. Szklarczyk et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, 39, 2011. [70](#)
- K. Thulasiraman and M. N. S. Swamy. *Graphs: Theory and Applications*. John Wiley and Son, 1992. [24](#), [25](#)
- B. P. Tu, A. Kudlicki, M. Rowicka, and S. L. McKnight. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science*, 310:1152–1158, 2005a. [xx](#), [132](#), [133](#), [135](#)
- B. P. Tu, A. Kudlicki, M. Rowicka, and S. L. McKnight. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science*, 310:1152–1158, 2005b. [138](#), [174](#)
- T. M. Venancio, S. Balaji, and L. M. Iyer. .reconstructing the ubiquitin network - cross-talk with other systems and identification of novel functions. *Genome Biology*, 10:R33, 2009. [106](#)
- P.K. Vohra, V. Puri, T.J. Kottom, A.H. Limper, and C.F. Thomas. Pneumocystis carinii ste11, an hmg-box protein, is phosphorylated by the mitogen activated protein kinase pcm. *Gene*, 312:173–179, 2003. [xxiv](#), [157](#), [158](#)
- A.I. Volpert and S.I. Hudjaev. *Analysis in classes of discontinuous functions and equations of mathematical physics, Chapter 12*. Marinus Nijhoff Publishers, Netherland, 1985. [34](#), [35](#), [49](#), [50](#)

## REFERENCES

---

- C. von Mering, R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417:399–403, 2002. [66](#)
- J. Warringer and A. Blomberg. Evolutionary constraints on yeast protein size. *BMC Evol. Biol.*, 6(61), 2006. [2](#)
- M. G. Wilkinson, M. Samuels, T. Takeda, W. M. Toone, J. C. Shieh, T. Toda, J. B. A. Millar, and N. Jones. The Atf1 transcription factor is a target for the Styl stress-activated MAP kinase pathway in fission yeast. *Genes*, 10:2289–2301, 1996. [97](#), [98](#)
- D. H. Wolpert. Determining whether two data sets are from the same distribution. In *15th international workshop on Maximum Entropy and Bayesian Methods*, 1995. [90](#)
- E. Yeager-Lotem et al. Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *PNAS*, 101: 5934–5939, 2004. [4](#), [12](#), [55](#), [58](#)
- S. H. Yook, Z.N. Oltvai, and A. L. Barabasi. Functional and topological characterisation of protein interaction network. *Proteomics*, 4(4):928–942, 2004. [4](#)
- H. Yu and M. Gerstein. Genomic analysis of the hierarchical structure of regulatory networks. *PNAS*, 103:14724–14730, 2006. [xvi](#), [54](#), [55](#), [60](#), [65](#), [75](#), [108](#), [145](#)
- F. T. Zenke, W. Zachariae, L. Lunkes, and K. D. Breunig. Gal80 proteins of *kluyveromyces lactis* and *Saccharomyces cerevisiae* are highly conserved but contribute differently to glucose repression of the galactose regulon. *Molecular and Cellular Biology*, 13:7566–7576, 1993. [101](#)

## REFERENCES

---

S. Zenker, J. Rubin, and G. Clermont. From inverse problems in mathematical physiology to quantitative differential diagnoses. *PLOS Computational Biology*, 3:2072–2086, 2007. [141](#)